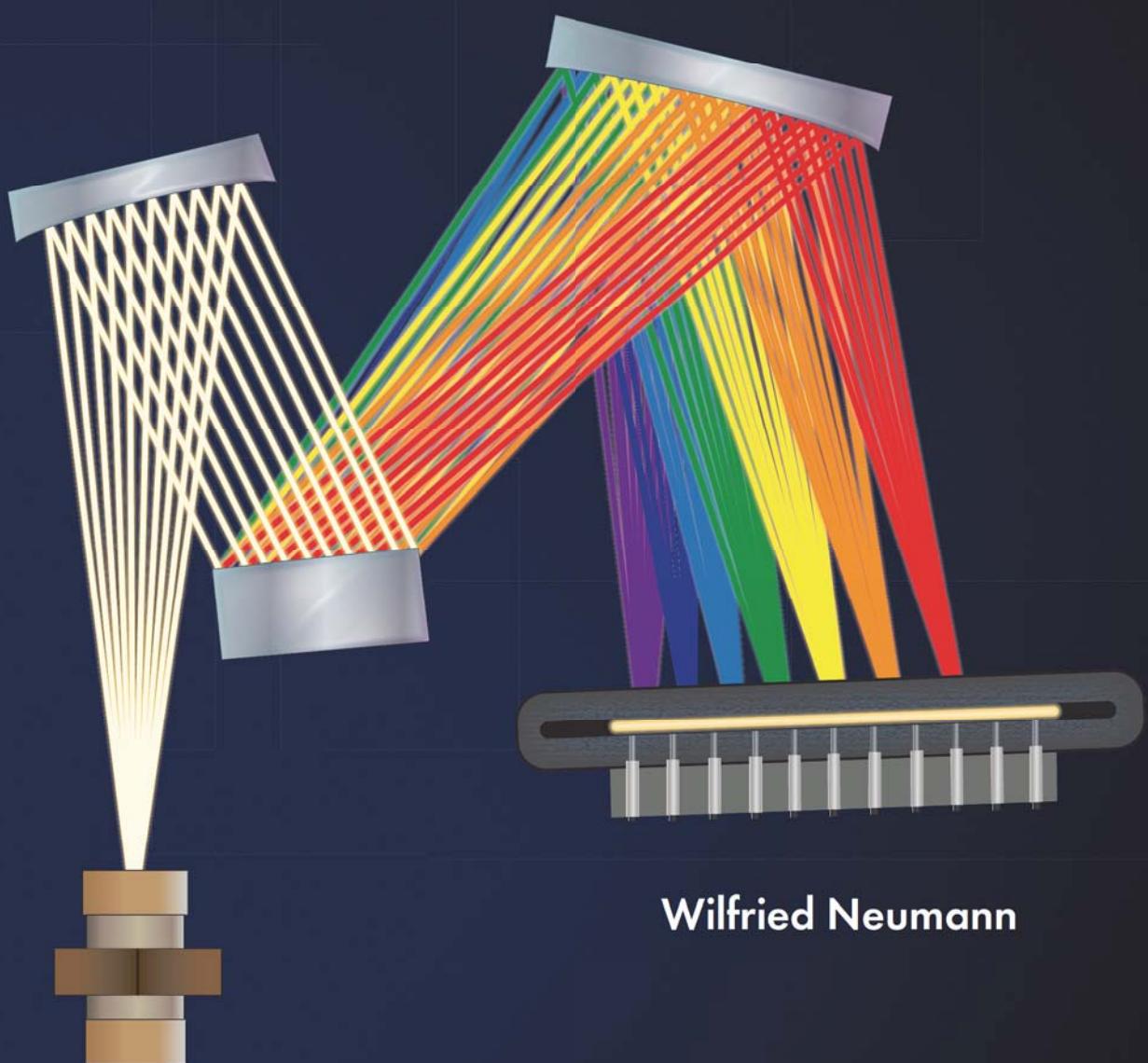


SPIE.

Fundamentals of Dispersive Optical Spectroscopy Systems



Wilfried Neumann

Fundamentals of
**Dispersive Optical
Spectroscopy Systems**

Fundamentals of
**Dispersive Optical
Spectroscopy Systems**

Wilfried Neumann

SPIE PRESS
Bellingham, Washington USA

Library of Congress Cataloging-in-Publication Data

Neumann, Wilfried, author.

Fundamentals of dispersive optical spectroscopy systems/Wilfried Neumann
pages cm

Includes bibliographical references and index.

ISBN 978-0-8194-9824-3

1. Optical spectroscopy. 2. Optical spectrometers. I. Title.

QC454.O66N48 2014

535.8'4—dc23

2013032275

Published by

SPIE—The International Society for Optical Engineering

P.O. Box 10

Bellingham, Washington 98227-0010 USA

Phone: +1 360 676 3290

Fax: +1 360 647 1445

Email: spe@spie.org

Web: <http://spie.org>

Copyright © 2014 Society of Photo-Optical Instrumentation Engineers (SPIE)

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.

The content of this book reflects the work and thought of the author(s). Every effort has been made to publish reliable and accurate information herein, but the publisher is not responsible for the validity of the information or for any outcomes resulting from reliance thereon.

Printed in the United States of America.

Second printing

SPIE.

Table of Contents

<i>Preface</i>	<i>xiii</i>
<i>Glossary of Symbols and Notation</i>	<i>xv</i>
1 Introduction, Terminology, and Scales	1
1.1 General Introduction	1
1.2 Photon Energies	2
1.3 Photon-Energy Conversion Equations	3
1.4 Naming Convention	4
1.5 The Spectral Line	5
1.6 General Rule of Optical Transfer	5
1.7 Definitions	6
1.7.1 Exponential functions and signal damping (attenuation)	6
1.7.2 Low-pass filter functions	8
1.7.2.1 A note on the dB interpretation	9
1.7.3 Definition of bandwidth in electric versus optical spectroscopy systems	10
1.8 Spectral Distribution of Thermal Radiation by Planck's Law	10
1.9 Keeping Optics Clean	11
References	11
2 Spectrometer Concepts	13
2.1 Basic Principle of an Optical Spectrometer	13
2.1.1 Attributes of modular spectrometers	14
2.2 Basic Grating Parameters and Functions	15
2.2.1 The free spectral range	16
2.2.2 Dispersion of gratings and prisms	16
2.3 Existing Basic-Spectrometer Concepts	17
2.3.1 The Littrow configuration	17
2.3.2 Ebert–Fastie configuration	20
2.3.2.1 Origin of astigmatism	23
2.3.3 Czerny–Turner configuration	23
2.4 Impacts and Distortions to Spectrometers	24
2.4.1 The influence of the internal angles on the wavelength	25
2.5 Other Spectrometers, Including Those for the Vacuum Range	25

2.5.1	Curved-grating spectrometer: Wadsworth setup	27
2.5.2	Normal incidence	27
2.5.3	Seya–Namicka	27
2.5.4	Grazing incidence	27
2.5.5	Rowland circle spectrometer and Paschen–Runge mount	28
2.6	Other Parameters and Design Features	28
2.6.1	Straight slits versus curved slits	28
2.6.2	Aperture and light flux (luminosity)	29
2.6.2.1	Real aperture or <i>f</i> -number?	29
2.6.2.2	Examples of the influence of the internal angles on the light flux	30
2.6.2.3	Calculating the <i>f</i> -number versus Ω for light flux/luminosity	31
2.6.3	Dispersion of spectrometers	36
2.6.4	Intensity distribution in the exit	37
2.6.5	Spectral resolution	39
2.6.5.1	Does one measure the resolution of the spectrometer or that of the experiment?	42
2.6.5.2	When is a spectral curve completely reproduced?	43
2.6.5.3	Rayleigh diffraction limit	44
2.6.5.4	Resolution of a monochromator compared to that of a spectrograph	45
2.6.5.5	Numerical resolution R_p and R_r , and their wavelength dependence	48
2.6.6	Image quality: Q-factor or fidelity	51
2.6.6.1	Calculating aberrations	51
2.6.6.2	A reflecting spectrometer has two or even three axes	52
2.6.6.3	The slit height also influences the total aberrations	52
2.6.6.4	The quality factor	53
2.6.6.5	Image-transfer issues	53
2.6.6.6	Spectrometers with internal image correction	56
2.6.6.7	General aberrations and coma	57
2.6.7	False and stray light, and contrast	58
2.6.7.1	Reducing stray light	59
2.6.8	Contrast ratio C	60
2.7	Mechanical Stability and Thermal Influence	60
2.7.1	Measuring thermal variations	61
2.7.2	Defocusing effects	61
2.7.3	Typical thermal constants	61
2.7.4	Minimizing environmental influence	63
2.8	Reduction of Unwanted Spectral Orders, and Other Filtering	63
2.8.1	Long-pass filters	63

2.8.2	Band-pass filters and prism	64
2.8.3	Short-pass filters	65
2.8.4	General filtering techniques	66
2.8.5	Notch filters	66
2.9	General Collection of Performance Parameters of Spectrometers	66
	Reference	67
3	The Dispersion Elements: Diffraction Grating and Refraction Prism	69
3.1	Introduction	69
3.2	Diffraction Efficiencies and Polarization of Standard Gratings	69
3.3	Types of Dispersers	71
3.3.1	Holographic gratings	71
3.3.2	Echelle gratings	72
3.3.3	Concave and other curved gratings	73
3.3.4	Transmission gratings	73
3.4	The Prism	73
3.5	The Grism	75
3.6	Other Features of Diffraction Gratings	77
3.6.1	Polarization anomaly	77
3.6.2	Polarization of Echelle gratings	78
3.6.3	Scattering effects	79
3.6.4	Grating ghosts	79
3.6.5	Shadowing and diffusion	80
3.6.6	Surface coating	80
	References	81
4	Design Considerations of Monochromator and Spectrograph Systems	83
4.1	Beam Travel inside a Spectrometer	83
4.1.1	Beam travel in a symmetric spectrometer	85
4.1.2	Variations of the basic Ebert–Fastie and Czerny–Turner concepts	87
4.1.3	Output wavelength as a function of the source position	88
4.1.4	Local output dispersion as a function of the lateral position in the field output	89
4.1.5	Output dispersion and fidelity as a function of the tilt angle of the field output	90
4.1.6	Correction methods for spectral imaging	92
4.1.6.1	External imaging correction	92
4.1.6.2	Internal imaging correction by toroidal mirrors	93
4.1.6.3	Internal imaging correction by a curved grating	94
4.1.6.4	Internal imaging correction by a Schmidt corrector	94
4.1.7	Prism spectrometer	95
4.1.8	Dispersion of a prism spectrometer	96
4.1.9	Echelle grating spectrometers	99
4.1.10	Transmission spectrometers	99

4.2	Grating Rotation and Actuation	100
4.2.1	Classical driving system	101
4.2.2	Grating actuation by a rotating system	102
4.3	Multiple-Stage Spectrometers	105
4.3.1	Double-pass spectrometers	106
4.3.2	Double spectrometers	107
4.3.2.1	Subtractive spectrometers	108
4.3.2.2	Efficiency behavior and analysis	110
4.3.2.3	Energy transmission and bandwidth of single-, double-, and triple-stage spectrometers	110
4.3.2.4	Effects of photon traveling time (time of flight)	111
4.3.3	Construction considerations for double spectrometers	112
4.3.3.1	Additive setup	112
4.3.3.2	Subtractive setup	113
4.3.3.3	Modern off-axis double spectrometers	114
4.3.3.4	Mechanical filtering in double spectrometers	118
4.3.4	Various configurations of flexible double spectrometers	118
4.3.5	General performance data of double spectrometers versus similar single-stage systems	119
4.3.6	Triple-stage spectrometers	120
4.4	Echelle Spectrometers	122
4.4.1	Echelle monochromators and 1D spectrographs	123
4.4.2	High-resolution Echelle spectrometer designed as a monochromator and 1D spectrograph	124
4.4.2.1	Echelle aberrations	127
4.4.2.2	Thermal drift assuming an aluminum chassis	127
4.4.3	Two-dimensional Echelle spectrometer for the parallel recovery of wide wavelength ranges at high resolution	128
4.4.3.1	Concept of a compact 2D Echelle	131
4.4.3.2	Comparison of an Ebert–Fastie and a folded Czerny–Turner	133
4.4.3.3	Constructive precautions	136
4.5	Hyperspectral Imaging	137
4.5.1	Internal references	137
4.5.2	Example of hyperspectral imaging	137
4.5.2.1	Image reproduction and spectral recovery	139
4.5.2.2	Overlaid hyperspectral image recovery	139
4.5.2.3	Separated hyperspectral image recording	140
4.5.2.4	Hyperspectral imaging supported by filters	140
4.5.3	General design for hyperspectral imaging	141
4.5.3.1	Design considerations	141
	References	142

5 Detectors for Optical Spectroscopy	143
5.1 Introduction	143
5.1.1 Work and power of light signals	143
5.1.2 Basic parameters of detectors	143
5.1.2.1 Pre-amplifier considerations and wiring	144
5.1.2.2 General signals and sources of noise in optical detector systems	145
5.1.3 Detection limit, noise, and SNR	146
5.1.4 Detection limit, noise, and SNR in absolute measurements	146
5.1.5 Detection limit, noise, and SNR in relative measurements	147
5.2 Single-Point Detectors	147
5.2.1 Phototubes	147
5.2.2 Comments on the interpretation of PMT data sheets	150
5.2.3 A sample calculation for PMTs, valid for an integration time of 1 s	150
5.2.4 Photon counter	150
5.2.5 UV PMTs and scintillators	151
5.3 Illumination of Detectors, Combined with Image Conversion	152
5.4 Channeltron® and Microchannel Plate	153
5.5 Intensified PMT and Single-Photon Counting	157
5.6 Solid State Detectors	157
5.6.1 General effect of cooling	159
5.6.2 Planck's radiation equals blackbody radiation	159
5.6.3 Detectors and the ambient temperature	160
5.6.3.1 Signal modulation and synchronized detection	162
5.6.3.2 Estimation of the modulated measurement	168
5.6.4 Tandem detectors	169
5.6.5 Typical parameters of solid state detectors, and their interpretation	170
5.7 Design Considerations of Solid State Detectors	171
5.7.1 Illumination of small detector elements	171
5.7.2 Charge storage in semiconductor elements, thermal recombination, and holding time	171
5.7.3 PIN and avalanche diodes	172
5.7.4 Detector coupling by fiber optics	172
5.8 Area Detectors: CCDs and Arrays	173
5.8.1 Mounting of area detectors, the resulting disturbance, and the distribution of wavelengths	173
5.8.1.1 Popular versions of area detectors	174
5.8.2 Basic parameters of arrays and CCDs with and without cooling	175
5.8.2.1 Pixel size, capacity, sources of noise, dynamic range, shift times, read-out time, and ADC conversion time	176
5.8.2.2 Applicability of CCDs for spectroscopy, image processing, and photography	178

5.8.3	Signal transfer and read-out	178
5.8.3.1	Combining the read-out in imaging mode and the display in spectroscopy mode	181
5.8.4	CCD architectures	181
5.8.5	CCD and array efficiency	183
5.8.5.1	Front-illuminated CCDs	183
5.8.5.2	Rear-side-illuminated CCDs	183
5.8.5.3	Interference of rear-side-illuminated CCDs: Etaloning	186
5.8.6	Time control: synchronization, shutter, and gating	187
5.8.6.1	Shutter control	187
5.8.6.2	Microchannel-plate image intensifiers	188
5.8.7	Current formats of area detectors	188
5.8.8	Read-out techniques: Multi-spectra spectroscopy, binning, and virtual CCD partition	189
5.8.8.1	Virtual CCD programming	192
5.8.9	CCDs and array systems with image intensification	193
5.8.9.1	CCDs with on-chip multiplication or electron multiplication (EMCCD)	193
5.8.9.2	CCDs with an additional microchannel-plate image intensifier (MCP-CCD)	194
5.8.10	Data acquisition in the ms– μ s time frame	195
5.8.10.1	Kinetic measurements	195
5.8.10.2	Double-pulse measurements	197
5.8.11	Extending the spectral efficiency into the deep UV	198
5.8.12	NIR and IR area detectors	198
5.9	Other Area Detectors	200
5.9.1	CID and CMOS arrays	200
5.9.1.1	Typical CMOS parameters, and comparison to CCDs	200
5.9.2	Position-sensitive detector plate	203
5.9.3	Streak and framing camera	203
	References	205
6	Illumination of Spectrometers and Samples: Light Sources, Transfer Systems, and Fiber Optics	207
6.1	Introduction and Representation of Symbols	207
6.2	Radiometric Parameters	208
6.3	Advantage of Using Ω and sr	210
6.4	Different Types of Radiation and Their Collection	210
6.4.1	Laser radiation	210
6.4.2	Cone-shaped radiation	213
6.4.3	Ball-shaped radiation from point sources: Lamps	215
6.4.3.1	Thermal filament lamps	216
6.4.3.2	Arc discharge lamps	216

6.4.3.3	Spectra of the various lamp types	217
6.4.3.4	Light collection and transfer into a spectrometer	218
6.4.4	Diffuse radiation collected by integrating spheres	219
6.4.4.1	Collecting lamp radiation	222
6.4.4.2	Approaching the parameters of a sphere	222
6.4.5	NIR radiation	224
6.4.6	IR radiators	225
6.5	Examples of Optimizing Spectrometer Systems	227
6.5.1	Optimization of gratings	227
6.5.2	Change-over wavelengths of lamps, gratings, and detectors	228
6.6	End Result of an Illumination Monochromator System	230
6.7	Light Transfer and Coupling by Fiber Optics	231
6.7.1	Fiber guides, light-wave guides, and fiber optics	231
6.7.2	Fiber optics for the UV–Vis–NIR range	232
6.7.3	Fiber optic parameters and effects	233
6.7.4	“Flexible optical bench,” and a precaution about its handling	236
6.7.5	Typical kinds and variations of single fibers and fiber cables	236
6.7.5.1	Basic versions	236
6.8	Transfer Systems	239
6.8.1	Coupling by bare optical fibers	239
6.8.2	Coupling by lens systems	240
6.8.3	Coupling by mirror systems	242
	References	243
7	Calibration of Spectrometers	245
7.1	Calibration of the Axis of Dispersion, Wavelength, and Photon Energy	245
7.1.1	Parameters that define the angular position of a dispersion element	245
7.1.2	Driving a grating or prism spectrometer	245
7.1.2.1	Grating spectrometers with a sine-functional drive	246
7.1.2.2	Calibrating a scanning system with a sine drive	247
7.1.2.3	First calibration of a sine-driven system	247
7.1.2.4	Parameters that can degrade the linearity	248
7.1.2.5	Timing calibration checks and recalibration	248
7.1.2.6	Recalibration requirements	249
7.1.3	Grating spectrometers with a rotary drive	249
7.1.4	Calibration of the field output	251
7.1.4.1	Output dispersion as a function of the lateral position in the field output	252
7.2	Calibrating the Axis of Intensity, Signal, and Illumination	253
7.2.1	Requirements for a useful calibration and portability of data	253
7.2.2	Light sources for radiometric calibration	253
7.2.3	Procedures to produce reliable calibrated data	254

7.3	Transfer Efficiency of Spectrometers	256
7.3.1	General behavior	256
7.3.2	Measurement of transfer efficiency	256
	References	258
8	Stray and False Light: Origin, Impact, and Analysis	259
8.1	Origin of Stray Light	259
8.2	Impact of Stray Light	261
8.2.1	Disturbance in the application of discrete spectral signals	261
8.2.2	Disturbance in the application of broadband spectral signals	263
8.3	Analysis and Quantization of Stray Light in Spectrometers and Spectrophotometers	264
8.4	Minimizing the Impact of Disturbance through Optimization	266
8.5	Reducing Stray Light	267
	References	268
9	Related Techniques	269
9.1	Compact, Fiber-Optic-Coupled Spectrographs	269
9.2	Programmable Gratings	272
9.3	Bragg Gratings and Filters	272
9.4	Hadamard Spectrometer	273
9.4.1	Principle of Hadamard measurements	274
9.4.2	Hadamard setups	274
	References	275
	<i>Index</i>	277

Preface

My search for universal and comprehensive literature on dispersive optical spectroscopy revealed many gaps. The books on very basic information are rather theoretical and dig deep into arithmetic derivations to calculate spectrometers, illumination, and detection. The books on the different applications of optical spectroscopy are mainly “cookbooks” and do not explain why something should be done in a certain way. Books with comprehensive content are available from the vendors of dispersers, spectrometers, detectors, and systems—they naturally feature the advantages of the supported products but offer no overall view.

For more than twenty years, I have calculated and delivered special dispersive spectroscopy systems for different applications. In the time between inquiry and decision, the customers wanted to justify my presentation and compare it. A common problem was finding useful references that could be used to verify my calculations and predictions. So, again and again, I wrote long letters combining the different parameters of the project presented. Several of my customers—industrial project managers as well as researchers—not only acknowledged the proposals but also often used the papers to check the instrumental performance at delivery. Because the proposals fit the requirements and the predictions were at least reached, their confidence was earned. Customers used my papers for internal documentation and teaching. Several asked me to provide the know-how in a general, written database in order to close the gap between theory, practice, and applications. After my retirement from regular work, I did just that, and published my writing on my homepage (www.spectra-magic.de). Now, the content has been improved and extended into a pair of printed books, the first of which you are reading now.

The aim of this book is to supply students, scientists, and technicians entering the field of optical spectroscopy with a complete and comprehensive tutorial; to offer background knowledge, overview, and calculation details to system designers for reference purpose; and to provide an easy-to-read compendium for specialists familiar with the details of optical spectroscopy.

Acknowledgments

My thanks are first addressed to my wife, Heidi, for her patience during the months spent investigating, reviewing, and writing. I also thank those who urged me to start writing in the first place and who collected data and calculations. The trigger to turn the homepage into written books came from Dr. Karl-Friedrich Klein, who kept me going and contacted SPIE. The section on fiber optics was supported by Joachim Mannhardt, who provided specifics and added features and ideas. After the manuscript was given to SPIE, external reviewers spent much effort on the content, providing corrections and suggestions for improvement; that valuable support came from Mr. Robert Jarratt and Dr. Alexander Scheeline. Last but not least, I'd like to thank Tim Lamkins, Scott McNeill, and Kerry McManus Eastwood at SPIE for the work they invested into the project.

I hope that readers will find useful details that further their interest or work.

Wilfried Neumann

May 2014

Glossary of Symbols and Notation

A	Absorbance (extinction) in photometric absorption measurements
A	Geometric area
A	Light angle inside a prism
ADC (A/D-C)	Analog-to-digital converter
A_{iG}	Effective disperser area at a given disperser angle
A_{iM}	Illuminated area of the focusing mirror
B	Bandwidth
C	Capacity
C	Contrast; ratio of useful signal/disturbance
c_0	Speed of light
CCD	Charge-coupled device
d	Deflection angle at the prism
d	Dispersed beam after a grating
D^*	Numeric capability of an IR detector for the recovery of low signals
dB	Decibel
d_x	Focus displacement after thermal change
d_y	Focus increase after thermal change
e	Base of the natural logarithm
E	Deformation factor at the exit of a spectrometer
e^-	Electron
$E_{(\lambda)}$	Irradiance of a light beam on a normalized surface
el	Elbow angle
eV	Electron volt
f	Focal length
f	Frequency
f_c	Angular frequency
FSR	Free spectral range
FWHM	Full width at half maximum
h	Planck's constant (6.626×10^{-34} Js)

<i>h</i>	Slit height
H	Total aberration
<i>hb</i>	Number of pixels binned together
I	Parallel incident beam to grating or prism
<i>i</i> ₁	Angle of the prism's incident light related to N
J	Joule
<i>k</i>	Absorption coefficient of a material
<i>k</i>	Boltzmann's constant (1.381×10^{-23} JK ⁻¹)
<i>k</i>	Grating constant for the distance of the grating lines
K	Kelvin
<i>K</i>	Thermal dilatation coefficient
<i>L</i>	Inductivity
<i>L</i>	Luminosity, light flux in spectrometers
<i>L</i> _(λ)	Radiance
LN	Liquid nitrogen
<i>m</i>	Modulation factor in lifetime measurements by phase/modulation
<i>m</i>	Spectral order number
<i>M</i>	Magnification factor
<i>M</i>	Radiant emittance/exitance
MCP	Microchannel plate; also, microchannel-plate image-intensifier system
<i>m</i> _s	Minimum slit width
<i>n</i>	f-number
<i>n</i>	Refractive index
<i>n</i>	Total number of lines in a grating
N	The normal of a grating or prism
O	Aberration
O ₁	Basic aberration
O _{ss}	Additive aberration
P	Power
PMT	Photomultiplier tube
PPS	Pulses per second; also, events per second
PSD	Phase-sensitive detector (in the lock-in); also, position- sensitive (counting) detector
<i>Q</i>	Energy of radiation <i>R</i> ; also, the numerical resolution
<i>Q</i>	Ratio of the numerical resolution <i>R_r/R_p</i>
<i>Q</i>	Quality factor
QE	Quantum efficiency
<i>r</i>	Radius of curved slits; also, the distance of the slit to the instrument's center
<i>R</i>	Normalized reflectance of a sample
<i>R</i>	Numeric resolution
<i>R</i>	Resistance

RD	Reciprocal dispersion
ROI	Region of interest
R_p	Theoretical resolution of a dispersing element
r_p	Absolute value of parallel polarization
r_s	Absolute values of perpendicular polarization
R_r	Real experimental resolution
s	Constant of thermal diffusion
SL	Number of vertical lines of a CCD
SNR (S/N-R)	Signal-to-noise ratio
sr	Steradian
SR	Number of horizontal register pixels of a CCD
STD	Standard deviation
T	Temperature; also, thermal change
T	Normalized transmission in photometric applications
w	Median distance of a mirror to the center line or grating center axis
W	Active grating or mirror width
W	Electrical or optical work
x	Geometric dilation as a function of thermal change
x	Half the inclusion angle at the grating
y	Geometric increase of the focal spot as a function of thermal change and dilatation
α	Angle of the light illuminating the grating or prism with respect to N
β	Angle of the diffracted or refracted light leaving the disperser with respect to N
δ	Inclusion angle of the light at the disperser originating from the lateral distance and width of the mirrors
δ	Phase angle or phase shift ellipsometry (SE)
Δ	Imaginary part of ellipsometric data
ε_1	Angle of the grating-impinging beam
ε_2	Angle of the beam leaving the grating
ι	Internal off-axis angle
ι_h	Horizontal off-axis angle in a spectrometer
ι_v	Vertical off-axis angle in a spectrometer
λ	Wavelength
ν	Oscillation frequency of a light wave
$\tilde{\nu}$	Frequency of a light wave presented as a wavenumber
ρ	Complex result of ellipsometric data
σ	Statistical parameter often used for deviations
τ	Time constant
Φ	Angle of sample illumination in ellipsometry
Φ	Median grating angle

Φ	Phase angle/phase shift in phase/modulation lifetime measurements
Φ	Radiant power/flux
Ψ	Real part of ellisometric data
ω	Angular frequency
ω	Normalized cone angle of illumination
Ω	Acceptance angle
Ω	Real and normalized aperture of a spectrometer; also, light-guiding factor

Chapter 1

Introduction, Terminology, and Scales

This book is intended to familiarize readers with the definition, design, justification, and verification of instrumentation for optical spectroscopy. Students and design engineers can use it as a background compendium. Experienced researchers and expert physicists will eventually find new interrelations of different parameters that are helpful for their work. Application scientists in biology and chemistry can use the book as a guideline to select the system best suited for their lab—eventually, purchasers will use it for the same reasons. The book does not deal with the basics of optical spectroscopy, however; it primarily concentrates on the application and technical realization of the technology. The physical basics of optical spectroscopy are well described by Ball.¹ Several books address light measurement in general, one of which is available from International Light.²

1.1 General Introduction

This book describes optical spectroscopy solutions, using dispersive spectrometers as the sole unit. In 99% of cases, a diffraction grating will be the heart of the spectrometer, creating a dispersion effect. The alternative element is the refractive prism, which is used in the minority of cases, while the combination of both is found in certain instruments. Because the grating is the most-used disperser, it assumes a central role throughout the book. After the introductory chapter, Chapter 2 describes different spectrometer concepts. The specifics of dispersion elements follow in Chapter 3, and Chapter 4 discusses practical spectrometer concepts available on the market. Chapter 5 brings up the different detector technologies in optical spectroscopy, while Chapter 6 deals with light sources, transfer systems, and illumination principles. To bridge the gap to applications, Chapter 7 deals with calibration methods, Chapter 8 discusses the impact of stray light

in general, and Chapter 9 provides an overview of related techniques. A collection of symbols and abbreviations, topics, figures, and equations is also provided. [This book's companion text, *Applications of Dispersive Optical Spectroscopy Systems* (to be printed by SPIE Press later in 2014), presents a collection of typical applications, their key parameters, functionality, and what to look for.]

Fundamentals of Dispersive Optical Spectroscopy Systems is not intended to guide designers to the very end of a development but rather help make basic decisions. For example, after a spectrometer design is chosen and the parameters are calculated, the optical path should be analyzed by a ray-tracing program—that will either help optimize the design or at least lead to possible improvements. The use of light source and detector information are similarly helpful, providing valid overall information, while details need to be gathered from the suppliers. Note that this book does not discuss compact mini-spectrometers: those are inaccessible, self-contained units that do not allow for any modification. Most often, compact systems are designed for a narrow range of applications, leaving only one alternative to the user: take it or look for a different solution. A short overview of compact spectrometers, among other techniques, is given in Chapter 9.

1.2 Photon Energies

The photons from different ranges of the electromagnetic spectrum (colors, wavelengths) carry different energies. Their behavior can be described by their wave nature, as electromagnetic vectors, or by their nature as quanta. In the case of waves, the frequency or wavelength is the measure of energy, and the orientation is described by the state of polarization. In the case of quanta, the energy is presented in joules (J), whereas the power is given in electron volts (eV). Both are directly related to the quantum character of a photon. Converting power or energy into waves will provide the wavelength λ (nanometers, nm), the frequency ν (gigahertz, GHz), or the wavenumber $\tilde{\nu}$ (cm^{-1}). Figure 1.1 presents the relation of the different scales used in optical spectroscopy.

The basic scale, showing the energy content (J), is rarely used. The other scales depend on the application. The wavelength λ is the most popular scale in applications in the ultraviolet (UV, 120–400 nm), visible (vis, 400–800 nm), and near-infrared (NIR, 800–3000 nm) range. For convenience, the wavelength can also be presented in microns ($1 \mu\text{m} = 10^3 \text{ nm}$), while small intervals of a nanometer are often written in picometers ($1 \text{ nm} = 10^3 \text{ pm}$). In literature before 1980, the Ångström (\AA) was as often used as the nanometer is today: $10 \text{ \AA} = 1 \text{ nm}$.

In the infrared (IR, 3–100 μm), the wavenumber $\tilde{\nu}$ is the most popular parameter. Special cases are scattering applications, such as Raman spectroscopy, where the wavenumber distance between signal and excitation

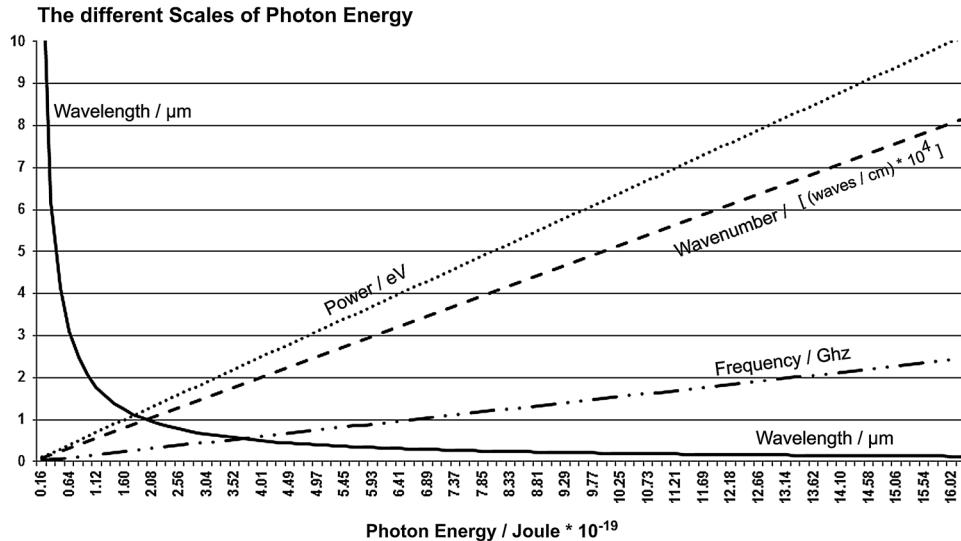


Figure 1.1 The different scales used in optical spectroscopy.

energy is presented. The reason is that the energy difference is constant and independent from the excitation energy. The frequency ν is most often used to specify the spectral bandwidth of laser signals and the resolution or bandwidth of interferometers. The electron volt is popular in material research and thin film applications, such as semiconductor spectroscopy or ellipsometry. In addition, it is the standard scale in deep-UV applications below 120 nm, requiring windowless, high-vacuum configurations. That technology is often used for solid state material research and treatment.

1.3 Photon-Energy Conversion Equations

All measures of energy are derived from the energy of photons, the electron, and the speed of light:

$$c_0 = 3 \times 10^8 \text{ m/s} \text{ [speed of light in vacuum or air (rounded)]}, \quad (1.1)$$

$$Q = h \times c_0 / \lambda = h\nu \text{ [J] (Basic energy of photons)}, \quad (1.2)$$

$$1 \text{ eV} = 1.602 \times 10^{-19} \text{ J (kinetic energy of an electron)}. \quad (1.3)$$

Via Q , the wavelength λ is connected to the frequency ν [Hz]:

$$\nu = c / \lambda \quad (1.4)$$

The conversion from wavelength λ to wavenumber $\tilde{\nu}$ [cm^{-1}]:

$$1 \text{ cm}^{-1} = (10^7 / \text{nm}) / 10. \quad (1.5)$$

A good approximation is

$$1 \mu\text{m} = 10,000 / \text{cm}^{-1}, \quad (1.6)$$

and vice versa. To convert from nm to eV,

$$1 \text{ eV} = (1,239,546.000/\text{nm})/1000. \quad (1.7)$$

An easy conversion is

$$1 \text{ eV} = 1240 \text{ nm}, \quad 1000 \text{ nm} (1 \mu\text{m}) = 0.8065 \text{ eV}. \quad (1.8)$$

The conversion from eV to wavenumber is

$$\text{eV} = \text{cm}^{-1}/8065. \quad (1.9)$$

The conversion between wavelength λ and frequency ν again uses the speed of light c_0 . Applying $\lambda = 1/c$, the according wavelength of 1 Hz (1/s) is $\lambda = 3 \times 10^8 \text{ m}$. With regard to the median range of optical spectroscopy, units nm (10^{-9} m) and GHz ($10^9 \times 1/\text{s}$) are preferred.

The conversion between frequency and wavelength is $\nu = c/\lambda$. For example, at 1 μm ,

$$\nu = (3 \times 10^8 \text{ m/s})/(10^{-6} \text{ m}) = 3 \times 10^{14}/\text{s} \quad \text{or } 3 \times 10^5 \text{ GHz},$$

which is abbreviated to 1 nm = 300/10⁶ GHz and 10⁶ GHz = 300/nm. To convert a known bandwidth from one scale to the other is very easy and is calculated by linear proportional rules, such as $\Delta\lambda/\lambda = \Delta\nu/\nu$ or $\Delta\text{eV}/\text{eV} = \Delta\tilde{\nu}/\tilde{\nu}$, and so on, generally expressed as

$$(\text{Bandwidth 1})/(\text{Median Value 1}) = (\text{Bandwidth 2})/(\text{Median Value 2}). \quad (1.10)$$

1.4 Naming Convention

The term “spectrometer” has two uses in the literature. The first use is to indicate an instrument that disperses or recombines light. It can operate as a “monochromator,” a “spectrograph,” or a “polychromator;” in other words, a building block. (These devices are discussed in further detail later and are the main topic of the book, besides selected periphery such as light sources and detectors.) The second meaning of “spectrometer” is used for a complete, application-oriented system, such as an “atomic emission spectrometer,” a “fluorescence spectrometer,” or an “absorption spectrometer.” All of these

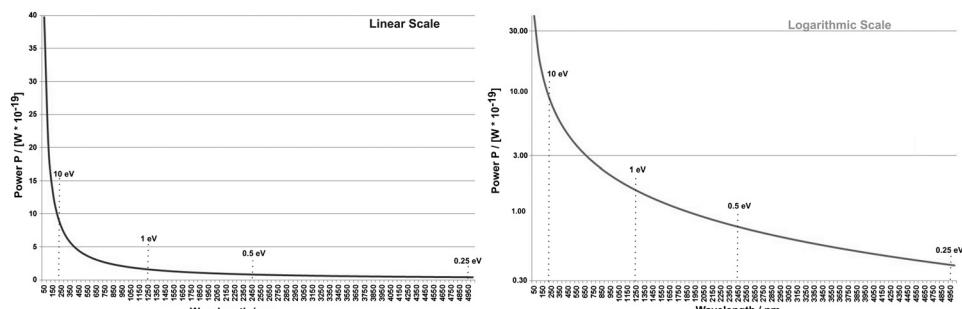


Figure 1.2 The electric power P in watts of a photon as a function of the wavelength.

systems are much better described as “spectrophotometers,” which are application-oriented systems that incorporate a spectrometer building block, some light-guiding equipment, and at least one detector. Thus, applications such as emission spectroscopy are not solved with a spectrometer but with an “emission spectrophotometer system.” This book always refers to the building block when the term “spectrometer” is used, whereas “spectrophotometers” and other application-oriented measurement systems are treated in the companion book.

1.5 The Spectral Line

Of course, an energetic spectral position is not 2D. The reason why discrete spectral information is called a “line” comes from history and from its appearance. The expression was coined by the first scientists in spectroscopy, who noticed that when a grating or prism is fully illuminated, the dispersed light of a certain wavelength appears line-shaped.

1.6 General Rule of Optical Transfer

As can be seen in Fig. 1.3, f/W is the aperture n :

$$M = (f_2/W_2)/(f_1/W_1), \quad (1.11)$$

where M is the magnification ratio, f is the focal length between object or image and optical component, and W is the active width of the optical component.

Equation (1.11) can also be written as $M = n_2/n_1$. O_1 and O_2 are the critical optical components; they can be lenses or mirrors. The reproduced size P_2 of an object (or focal spot) P_1 can also be calculated directly by $f_1/P_1 = f_2/P_2$, leading to M . iO represents other internal optics between the critical optics, such as

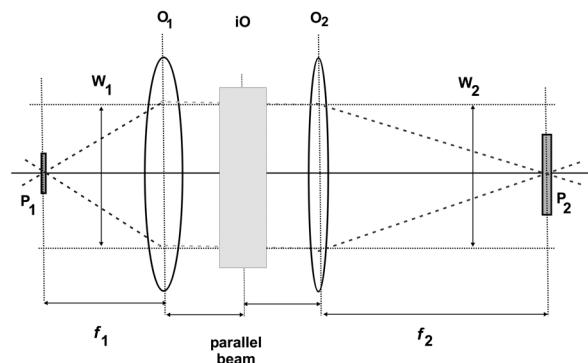


Figure 1.3 The relations between focal length f , used width W , and the reproduction size of an image P . Note that O is the optical component, lens, or mirror in the transfer path; iO is the inactive optical components (such as filters); and P is the object monitored by the system and the reproduced image of the object.

plane mirrors or gratings. So long as iO does not change the angles or focal lengths, they have no impact. In spectrometer applications, systems are often found sketched in external transfer optics, i.e., illumination optics. The impact of M will be found multiple times. Note that the rules only apply fully if the objects are in the focal spot.

1.7 Definitions

Throughout optical spectroscopy and signal recovery, some general functions are repeated. They define the dependence of response, attenuation, and bandwidth in both the energy and time domain.

1.7.1 Exponential functions and signal damping (attenuation)

The response of a system to the immediate change of the input signal typically follows an exponential function with the basis e (value 2.71828...). In spectroscopy, this is true for the response of detectors and electronics. The output change follows the “ e -function.”

Figure 1.4 depicts the behavior of e -functional processes that define the output response of a system upon the immediate change of the input signal. The rise or fall occurs in a flowing function, in which for every time constant

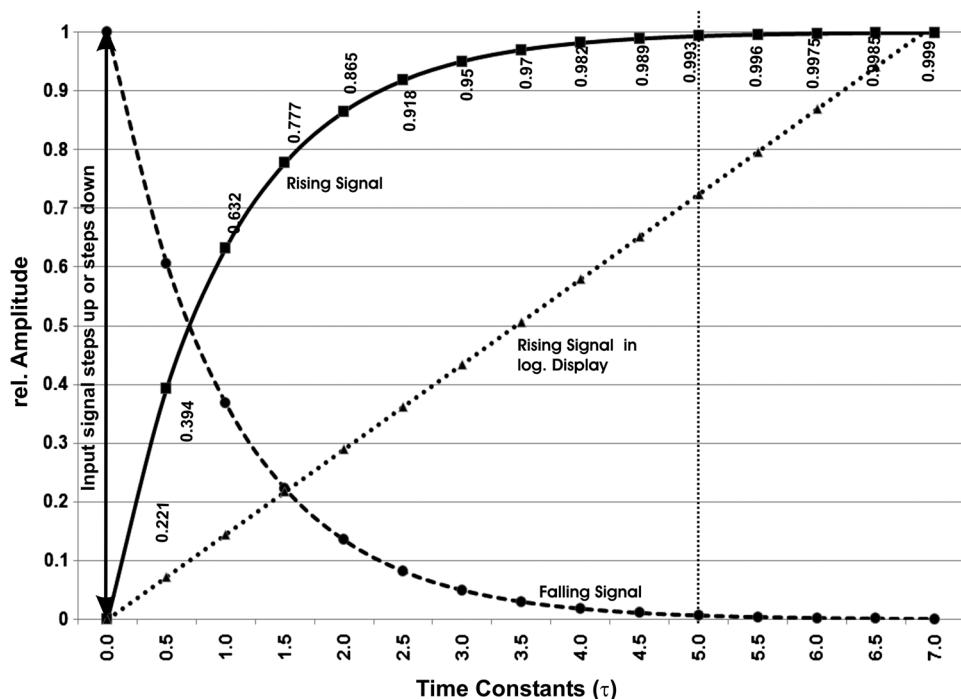


Figure 1.4 The general behavior of e -functional processes.

the final value is reached by steps of 63.21% from the remaining difference. After five time constants (5τ) and 99.326% of the original change, the function is considered to be accomplished or “swung-in.”

The simplest device to damp unmodulated signals is the combination of a resistor R with a charging capacitor C , or a resistor R in series with a coil L . The calculation of time constants is easy:

$$\tau = RC \quad \text{and} \quad \tau = L/R, \quad (1.12)$$

where R is in ohms (Ω), C is in farads (F), and L is in henries (H). Also important are voltage in volts and current in amperes.

- Based on the fact that $F = C/V$, with the dimensions As/V, the result is $\tau = \Omega \times (1/\Omega) \times s$, where s is seconds.
- By $\tau = L/R$ and $H = Vs/A$, the result is $\tau = \Omega s$.
- An RC circuit with $R = 10^6 \Omega$ and $C = 10^{-6}$ F results in $\tau = 1$ s.
- An LR series circuit with $L = 10^{-6}$ H and $R = 10^6 \Omega$ results in $\tau = \{1 s / (10^{-6} \times 10^6)\} = 1$ s.

If such a device receives sine wave signals, all frequencies will be more or less affected. The angular frequency of a system is required for the calculation of transfer functions:

$$\omega = 2\pi f, \quad \text{or} \quad \omega = (2\pi)/t. \quad (1.13)$$

The frequency, which represents the border frequency, is called f_c :

$$f_c = 1/(2\pi\tau); \quad (1.14)$$

thus, a device with $\tau = 1s$ has a border frequency of $f_c = 0.16$ Hz. The attenuation of an RC or LR device at 6 dB/octave will result in a value of 86% at f_c and 63% at $2f_c$. For better signal flattening with quick response (high frequencies), two or more time-constant devices are often combined in series, which increases the attenuation (roll-off) to 12 dB/octave (86% at f_c but 35% at the second octave) without affecting the response time.

The left side of Fig. 1.5 shows the simplest damping systems possible. The “input signal” arrow simulates an immediate change of the entrance signal. The solid curve on the right is the output signal for 5τ . Analog systems with a single damping device either consist of a series of a coil L and a resistor R , or a resistor R and a charging capacitor C against ground (the figure shows the equation for both). Single-stage analog devices always follow that function; it would require a series of multiple devices to modify the output function. Digital systems, on the other hand, provide many options to vary the response behavior. Digital systems contain other passive and/or active electronic components, in addition to the L/R or L/C combination in the entrance. A simple digital version is shown by the thin, dashed-dotted curves. For demonstration, the entrance circuitry of a digital system is considered to have a very short τ compared to the integration time. The internal signal will take the shape of a ramp after the integrator is started. At the end, it will have the

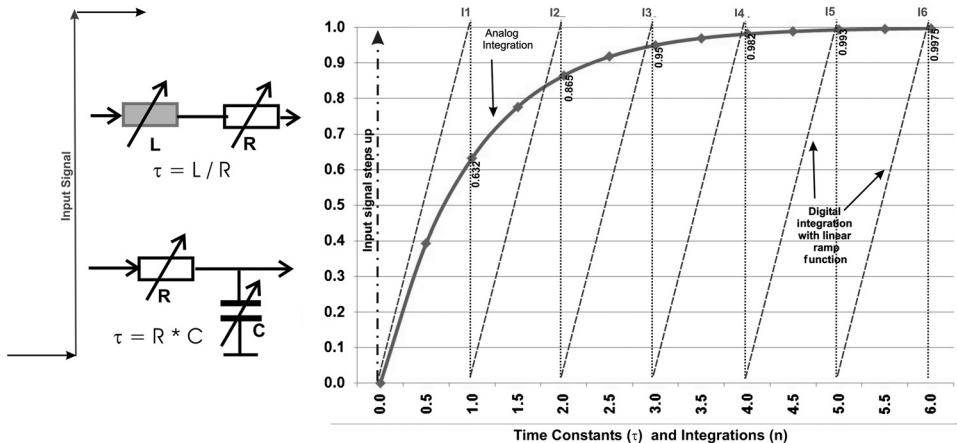


Figure 1.5 Definition of the time constant and comparison of integration methods.

final value (loaded with variance and noise, of course). The value will be digitized and stored, and then the next integration starts.

At the end of a series of integrations, the average value will be displayed and stored. In typical systems, the user will enter the integration time, and the number for one average (the appropriate τ) to fit the requirements will be set by the system. The data points stored will overlay (sliding average). The most versatile instruments incorporate digital signal processors (DSPs), which are able to change parameters “on the fly.” For instance, the integration can start with a very short time constant—for quick response—but as the ramp increases, the damping might also increase for better noise reduction.

In summary, it is not possible to compare DSP electronics with analog systems unless both use the same algorithms, in which case comparison makes sense. Otherwise, the quality of results versus measurement time will be the determiner. In optical spectroscopy, only single-point detectors (photomultipliers, photodiodes) can be read by pure analog electronics—all multiple detectors (array, CCD) work digitally. Thus, the integration, damping, and SNR behavior can differ strongly.

1.7.2 Low-pass filter functions

An important parameter in electronic measurements and optical spectroscopy is the filter function, especially that of the low-pass filter.

As shown in Fig. 1.6, the suppression of signals outside the desired bandwidth is defined in $-dB$ per octave. Precisely, in a low-pass filtered system, only the frequency of 0 Hz experiences no suppression. At the corner frequency f_c , the signal is already reduced to 86% of the original value, independent of the steepness (roll-off) of the filter function. The expression $-dB/oct$ describes the fact that for every frequency doubling, the suppression will drop with the same logarithmic output/input ratio. The graph is shown in linear plot. The dashed

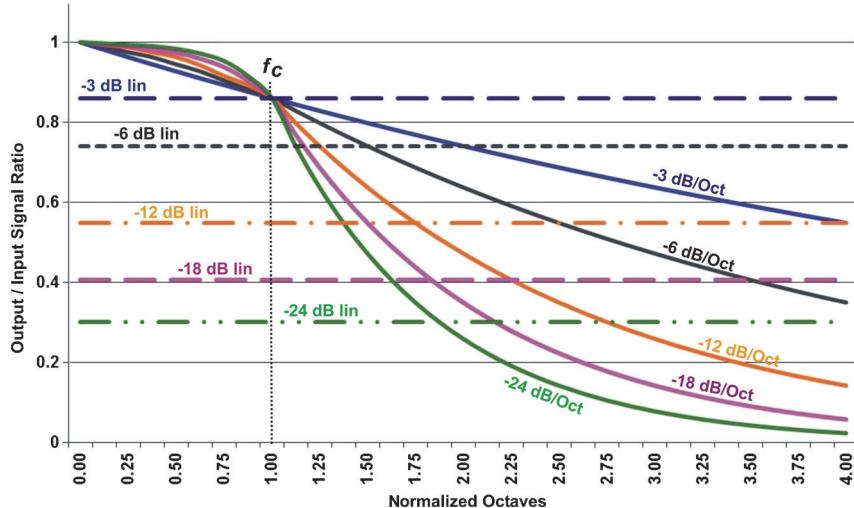


Figure 1.6 The suppression of signals outside the desired bandwidth.

lines mark the $-dB$ levels versus the normalized frequency. Note that with increasing $-dB/oct$ values, the damping at frequencies below f_c is weaker compared to smaller $-dB/oct$ values. A vertical mirror function, symmetric to the 0-Hz line, would display the behavior of a high-pass filter, and both curves combined would represent a bandfilter.

The low-pass is realized by a simple RC filter (such as that shown in Fig. 1.5). A simple, single-stage filter creates -6 dB/oct (Fig. 1.6). Steeper functions are generated by passing the signal through a series of RC circuitries, and each (equal) RC adds -6 dB/oct . Modern instruments include DSPs, which combine multiple filter functions with DC-amplification, providing widely improved flexibility and functionality. The steeper the roll-off function is, the flatter the top of the filter curve.

1.7.2.1 A note on the dB interpretation

Attenuation curves of signal curves and power curves differ from each other in calculation and steepness:

Signals, such as the output of detectors and amplifiers, are interpreted by the “voltage-level calculation,” the equation for which is

$$-dB = 20 \log_{10}(U/U_0). \quad (1.15)$$

The attenuation of -3 dB represents a factor of 0.86071 (“Voltage-dB”).

When dealing with “light power,” functions whose attenuation is calculated by the “power-level calculation” can be found according to the following equation:

$$-dB = 10 \log_{10}(P/P_0). \quad (1.16)$$

The attenuation of -3 dB represents a factor of 0.74082 (“Power-dB”).

The reason for the different calculation is that the voltage-level calculation is only based on one parameter, whereas the power-level calculation combines two (photons and watts), making it a squared function. In the equations, U and P are the output values of the system, and U_0 or P_0 are entrance values.

The filtering functions are fundamental for the behavior of detection systems (see Section 5.6.13).

1.7.3 Definition of bandwidth in electric versus optical spectroscopy systems

The bandwidth of a spectrometer is defined at half maximum of the signal (FWHM). In Fig. 1.7, this is shown by the solid lines. The integral under the area (representative for the power) within those limits is almost exactly 70% of the full area covered by the peak. For fast estimations, the upper part of the peak can be simplified to a triangle. In electrical calculations, the bandwidth is defined at maximum/ $\sqrt{2}$ or factor 0.7071 of the signal magnitude (dotted lines). The integral of that area is close to 50% of the full area covered by the peak.

1.8 Spectral Distribution of Thermal Radiation by Planck's Law

The effects and impacts created by thermal radiation not only create useful signals from light sources such as lamps, they are also responsible

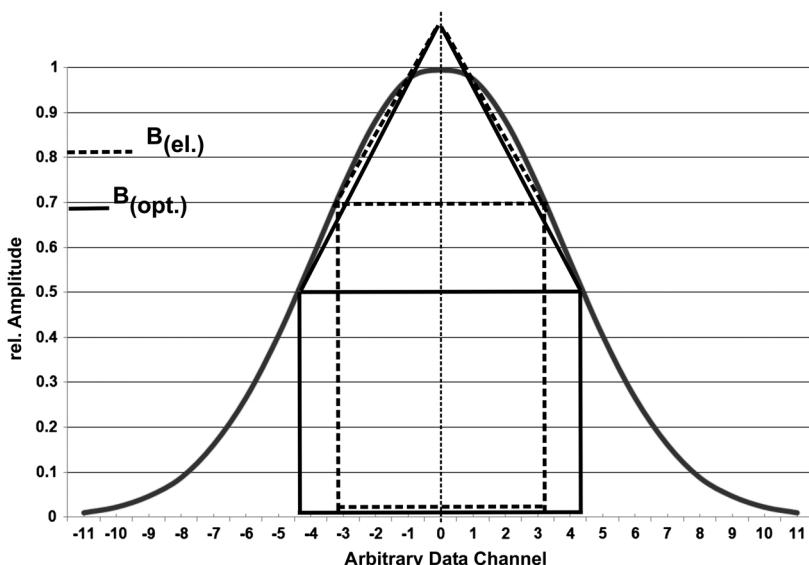


Figure 1.7 Comparison of the bandwidth definition for systems in electronic technology versus optical spectroscopy, with signal behavior following a Gaussian normal distribution.

for the major part of background signals and the associated noise. The law is

$$W_{e\lambda}[\lambda, T] = 8\pi h C \lambda^{-5} (e^{ch/k\lambda T} - 1)^{-1}, \quad (1.17)$$

where T is the absolute temperature of a blackbody in Kelvin (K), h is Planck's constant (6.626×10^{-34} Js), c is the speed of light (m/s), k is Boltzmann's constant (1.381×10^{-23} JK $^{-1}$), and λ is the wavelength (nm). The law and derivatives of it will be used several times throughout this book.

1.9 Keeping Optics Clean

In electronic systems, dirt and dust only become critical when cooling or other functional disturbance follows. In mechanical systems, functional impacts often become serious. In all optical systems, surfaces carrying dust, dirt layers, or invisible spills can create malfunctions to a certain extent, e.g., the specifications will drop, stray light will increase, wavelength-dependent behavior will change, etc. It is necessary to keep the optics clean, or, if an accident occurred, know how to remedy the situation.

Information on the general guidelines of cleaning optics is available elsewhere.³ Note that additional care is advised because mirrors and lenses are often coated. Thus, any handling treats the coating first and not the substrate. The manufacturers of spectrometers and optical parts should be able to offer advice about cleaning. Cleaning a prism is like cleaning a lens (or its coating). If a grating becomes dirty, there is only one thing to try: smoothly blowing clean, dry air or nitrogen at low pressure from the side over the surface can clean the surface. Any kind of touching or applying of liquids (even compressed air) will increase the damage of the surface. If the performance of the grating is not acceptable any more, it either needs to be replaced or sent back to the manufacturer for refurbishment.

References

1. D. W. Ball, *The Basics of Spectroscopy*, SPIE Press, Bellingham, WA (2001) [doi: 10.1117/3.422981].
2. A. Ryer, *Light Measurement Handbook*, International Light, Peabody, MA (1997).
3. R. Schalck, *The Proper Care of Optics: Cleaning, Handling, Storage, and Shipping*, SPIE Press, Bellingham, WA (2013) [doi: 10.1117/3.1001008].

Chapter 2

Spectrometer Concepts

2.1 Basic Principle of an Optical Spectrometer

A dispersive optical spectrometer (Fig. 2.1) requires five components. The **entrance aperture** is usually a slit but sometimes a pinhole or the end of a fiber. The purpose of the aperture is to limit the spatial distribution of the entering light and help define the light cone required to illuminate the system. Except in some rare cases, it is a basic requirement for a well-working spectrometer. For proper functioning, the light arriving at the aperture needs to be a clean cone of light.

The **collimating optics** can be a mirror or a lens; it ensures that the light will hit the **dispersing element** in a parallel, collimated fashion because gratings and prisms only work correctly if they are illuminated with collimated light. The optimal situation is reached when the collimator illuminates the disperser fully without sending light to the frame or passing by it. The dispersed light will leave the disperser in parallel bundles of light. Each wavelength leaves under a separate angle, but all light of the same wavelength and spectral order leaves under the same angle.

The dispersed light, traveling under a certain range of angles, will be collected by the **focusing optics**, again a mirror or a lens, which refocuses the spectrum in the output plane. At the **exit aperture**, there can be a slit, a pinhole, or a fiber. In the optimal case, the entrance aperture is re-imaged at the output. If the entrance was illuminated by a spectral interval finer than the system can resolve, the output light interval would be the same except for some losses or changes in polarization. The output plane can also be equipped with an area detector to collect the parallel arriving wavelengths.

If the optical components are perfectly spherical, the foci of the parallel output signals will arrive on the output radius of the focusing output element, i.e., the light will arrive in a curved field. The size Wi (diameter) of the optical components, the angles within the light path, and the quality of the components define the x/y size and quality of the field. Figure 2.1 shows a transmitting setup, which was chosen for easier description. In reality, reflecting grating systems are primarily used instead of transmitting systems

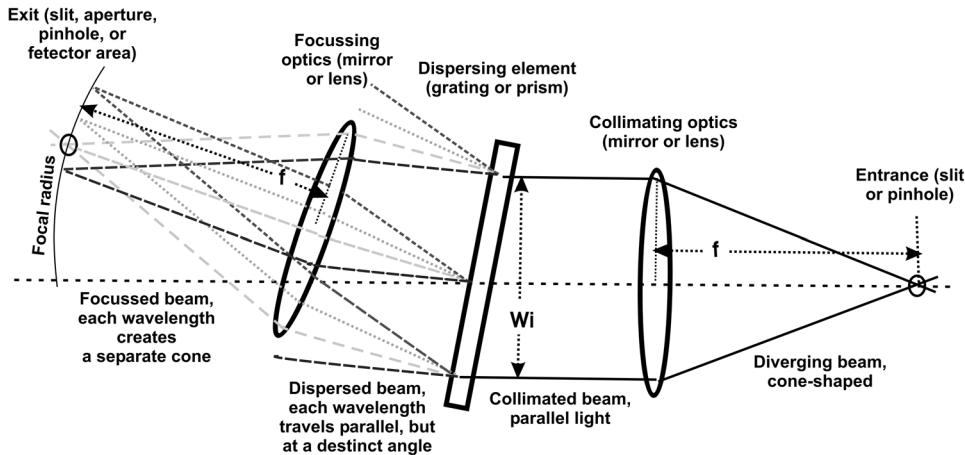


Figure 2.1 The components of a dispersive optical spectrometer.

because they provide a much wider range of use, more flexibility, and they are prevalent in laboratory use.

2.1.1 Attributes of modular spectrometers

- **Focal radius of the mirrors:** The longer radius f is, the better dispersion, resolution, and quality of spectral and image transfer, but also the lower the light flux (except in some special cases).
- **Size of the mirrors:** The bigger the diameter of the mirrors is, the higher the possible light flux. With the increase of the acceptance angles inside the system, the light-collecting power of the system increases, comparable with the lens diameter of a photographic system. The light collecting power $n = f/Wi$, where f is the focal length, and Wi is the used diameter (n is also called the f -number or aperture). Wider apertures and angles produce lower transfer quality.
- **The groove/line density (mm^{-1}) of the grating:** The higher the density is, the higher the dispersion and resolution, but also the lower the projected area at higher working angles of the grating. The width of the grating needs to be aligned with the width of the optics and the planned working angle. The larger the used width of the grating is, the better the light flux and resolution.
- **The size of entrance and exit aperture (normally a slit):** The wider the slit is, the more light enters (except when the beam itself is smaller than the slit), but the lower the resolution. The higher the slit is, the better the light flux at the expense of transfer quality, and therefore the resolution and image quality.
- **Internal angles:** The tighter the angles are, the better the geometric functions, leading to smaller components or increased focal length.

Knowledge of the grating's key parameters and their interrelation is required to assess the impact of the dispersing element on the functions of any spectrometer.

2.2 Basic Grating Parameters and Functions

Because the dispersing element plays a key role in all equations dealing with the dispersion and resolution of any spectrometer, it is necessary to consider a few disperser equations first. The grating dominates that field, so only those at the beginning are important.

Figure 2.2 illustrates the parameters of a reflecting diffraction grating. N is the grating normal, which is the perpendicular axis to the surface of the grating, and all other angles refer to N. The incident beam I hits the grating under angle α . Whatever angle α has, a share of its energy will be reflected. The angle of the surface reflectance sr is the mirror angle. The spectroscopist is concerned with only the diffracted part of the light that is accompanied by the reflected, absorbed, and scattered parts, all of which are discussed later. The main vector—the diffraction—is calculated by the **general grating equation**:

$$m \times \lambda = k \times (\sin \alpha \pm \sin \beta), \quad (2.1)$$

where m is the spectral order of diffraction; λ is the wavelength; k is the distance between the structures ($\text{inversion of } \text{mm}^{-1}$), also called the “grating constant;” I is the incident beam [the rays need to arrive in parallel (collimated) fashion]; d are the dispersed beams (they travel in parallel rays, but different wavelengths leave via the grating under different angles); α is the angle between the incidence beam I and N; $\pm\beta$ is the angle between the dispersed beam d and N (if α and β are on the same side of N, then $\alpha + \beta$, but if N is between α and β , then $\alpha - \beta$); and W is the ruled width of the grating (W/k equals the total number of lines available).

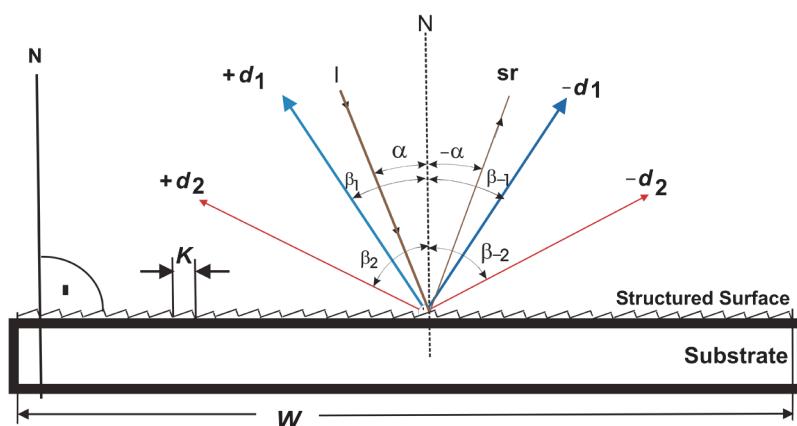


Figure 2.2 Parameters of a reflecting diffraction grating.

Consider two examples, according to Fig. 2.2. I has an angle α of 15 deg, and $+d_1$ has an angle β of 20 deg. k is 1 μm (equivalent to 1000 mm^{-1}):

$$m \times \lambda = 1 \mu\text{m} \times (0.259 + 0.342) = 0.601 \mu\text{m}.$$

If $m = 1$, the wavelength is 601 nm; if $m = 2$, the wavelength is 300.5 nm; etc. A recorded beam can contain multiple wavelengths and multiple orders that can overlay each other. For a given λ , there are an infinite number of β , all of which contain light. The spectrometer setup determines which d and/or β beam reaches the output. Figure 2.2 reveals more diffracted beams: $-d_1$ and $-d_2$, which are on the other side of N. For $-d_1$,

$$m \times \lambda = 1 \mu\text{m} \times (0.259 - 0.342) = (-1) \times 0.083 \mu\text{m},$$

which means that the first order will be at 83 nm. The negative sign indicates that the wavelength is on the other side of N and that the spectral orientation is inverted. (The impact on spectrometer functions is discussed in Sections 4.1 and 4.3.2.) Whether the sign is presented with the order or with the wavelength is unimportant; because negative wavelengths do not exist, it defines only the orientation.

If I is not a well-collimated beam, the grating will still work, but performance will drop.

2.2.1 The free spectral range

From Eq. (2.1), it is possible to derive the range that is available without spectral overlay, called the free spectral range (FSR):

$$\text{FSR} = \lambda/m. \quad (2.2)$$

A more precise interpretation is

$$\text{FSR} = \lambda_2 - \lambda_1, \text{ and } \lambda_2 = \lambda_1 + (\lambda_1/m), \quad (2.3)$$

where λ_1 is the lowest wavelength with noteworthy light intensity in the system, λ_2 is the highest wavelength without order overlay, and m is the spectral order used. Note that the first order needs the wavelength signals below λ_1 blocked if the FSR or λ_2 is exceeded. It does not need filtering of wavelengths above λ_2 . All higher orders need filtering of wavelengths outside the FSR in both wavelength directions (lower and higher).

2.2.2 Dispersion of gratings and prisms

The angles at the dispersion element are important for all dispersion calculations, as shown in Fig. 2.3. The basic calculation is performed by the angular dispersion

$$m \times d\lambda = k \times \cos \beta \times d\beta \quad (2.4)$$

or by

$$[d\beta/d\lambda] = [m/(k \times \cos \beta)], \quad (2.5)$$

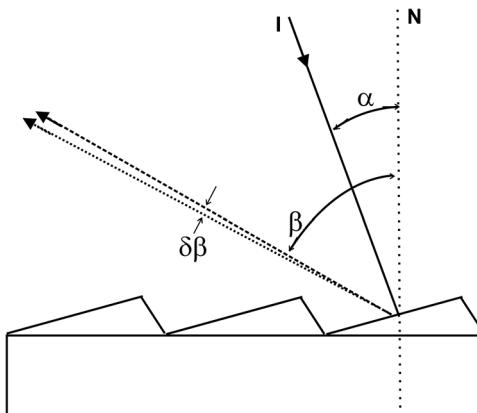


Figure 2.3 The angles at the dispersion element are important for dispersion calculations.

where α is the angle of illumination, β is the median angle of dispersion, $d\lambda$ is the small spectral difference within the dispersion angle, and $d\beta$ is the small angular difference within the dispersion angle.

Equations (2.4) and (2.5) are variations of Eq. (2.1): for example, consider the task of finding the angular change for the angular dispersion difference of the two mercury lines at 313.155 nm and 313.184 nm. There is a grating of 2400 mm^{-1} , and it is assumed that α and the median of β are identical. For 313.155 nm and by Eq. 2.1, $\alpha = \beta$ is found to be 22.07290076 deg . For 313.184 nm, the angle is $22.0750505237 \text{ deg}$, which is an angular difference $\delta\beta/\delta\lambda = 0.00214947 \text{ deg}/0.229 \text{ nm}$. Normalized, $\delta\beta$ is 0.0741 deg/nm .

The following sections proceed with a basic discussion of spectrometers based on these equations; dispersing elements are further treated in Chapter 3.

2.3. Existing Basic-Spectrometer Concepts

2.3.1 The Littrow configuration

When Otto Littrow presented his spectrometer in the 1860s (Fig. 2.4), it only consisted of three components: a pinhole, a prism, and a spherical mirror.

Besides transmitting systems, this setup is the most straightforward arrangement. It represents a theoretically ideal solution that realizes the basic requirements and has only three components: two very nearby slits, a mirror acting as the collimator and focusing unit, and a reflecting grating. In the original version, the Littrow grating has a hole in the center to let the incoming and outgoing light pass through, which leads to ideal angles, as shown in Fig. 2.5. The problem is that it is not very useful, which leads to solutions where the light beam opened in the horizontal or vertical axis to avoid the hole in the grating and the splitting of the entrance and exit. The grating was then moved aside, and the entrance and exit were also separated in the vertical plane, resulting in designs with small angles in the vertical and the horizontal direction.

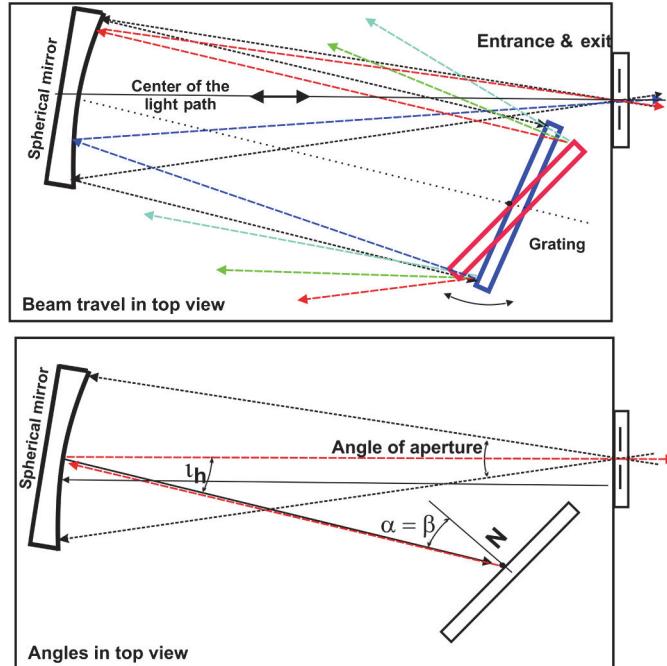


Figure 2.4 The Littrow mount seen from top.

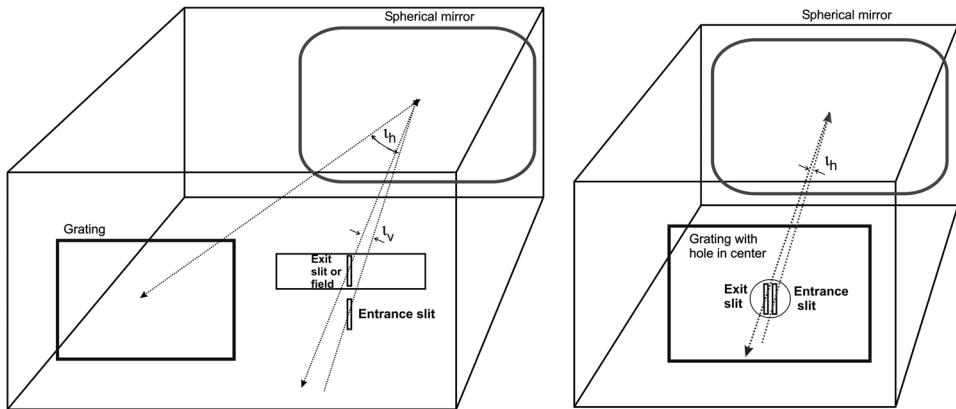


Figure 2.5 Littrow configuration in the spatial view.

Because the illumination and dispersion beams travel under the same angle, Eq. (2.1) is reduced to the Littrow equation:

$$m \times \lambda = k \times 2 \sin \alpha. \quad (2.6)$$

The surface efficiency of the reflecting mirror and disperser is found by Littrow surface correction:

$$A_{iG} = A_{iM} \times \cos \alpha \times \cos \iota_h \times \cos \iota_v, \quad (2.7)$$

where A_{iG} is the effective disperser area at a given disperser angle, A_{iM} is the illuminated area of the Littrow mirror, ι_h is the out-of-plane angle in the horizontal orientation, ι_v is the out-of-plane angle in the vertical orientation, and α is the illumination and dispersion angle.

The right image in Fig. 2.5 represents the original Littrow mount, with the pierced grating and very close slits. The mount provides almost ideal angles but no output field; it can only be used as a monochromator. Versions with only one slit plus a beamsplitter, and with two slits atop each other, have also been used. The left part shows a version used for many years as a monochromator and spectrograph. The monochromator version still carries the problem of having the entrance and exit very close together, allowing only small attachments to be mounted. However, in the spectrograph setup, it is very useful. Up to the early 1970s, such Littrows were used with photographic plates and films, most of which had very good optospectral parameters: at a typical $f = 2000$ mm, with 200-mm-wide ($f/10$) mirrors and grating, the out-of-plane angle ι is less than 5 deg. The out-of-plane angle is the opening angle between the beams running from slit to mirror and from mirror to grating. To achieve this, the mirror was only tilted 2.5 deg, requiring no curvature corrections, yet all rules of specular (direct) reflection still applied. If the illuminated detector field has a size with a 250-mm length and 25-mm height, the variation due to the curvature of the focal plane is only ± 0.75 mm. At a required spatial resolution of 100 line pairs (lp/mm or 10 μm) of the photographic plate, the displacement produces an image change from 10.0 to 10.3 μm , or 100 to 97 lp/mm, over the whole area. This scenario does not need correcting.

Littrows are not commercially available today, but they are still used in reference labs. For example, the characterization of gratings are almost always specified under the Littrow condition. If possible, Littrows are be used at $\alpha > \iota$ to avoid internal reflection and multiple dispersion. Both occur if the grating and mirror are approximately facing each other, sending the light back and forth. Because the angles α and β are identical, grating equation Eq. (2.1) is reduced to the Littrow equation [Eq. (2.6)].

In practice, all reflective systems suffer from the loss of projected surface coming from the out-of-plane angles in the illumination and dispersion arms. In the Littrow setup, the loss theoretically occurs in both directions, but because both follow the cosine function, the loss is small. A typical horizontal out-of-plane angle is 5 deg ($\cos 0.9962$), which is negligible. If the entrance and exit are 60 mm apart, thus allowing for a 25-mm entrance and exit slit height, the vertical ι becomes only 1.72 deg ($\cos 0.99955$). In a real instrument, both values play no role, but they are mentioned here for later comparison.

Note that the correction for the out-of-plane angles is not often found. However, it is shown in Figs. 2.4 and 2.7, and is essential for the calculation of surface efficiencies in reflecting systems. The area-correction equation is

$$A_{iG} = A_{iM} \times \cos \alpha \times \cos \iota_h \times \cos \iota_v,$$

where A_{iG} is the illuminated grating area; A_{iM} is the illuminated mirror area; α is the working angle of the grating, with respect to N; ι_h is the out-of-plane angle in the horizontal orientation; and ι_v is the out-of-plane angle in the vertical orientation.

To apply the values in the previous example, leave the grating at zero degrees, and thus $A_{iG} = A_{iM} \times 1 \times 0.9962 \times 0.9995$, which means that no correction is required. For all of the following spectrometer setups, the angles only open in one plane, making ι_v obsolete.

Advantages

- Superb image transfer, which produces extraordinary resolution, allows for high entrance slits and results in a large field at the exit, minimized internal angles, only one mirror, and a slim footprint.
- The optical setup only needs three degrees of adjustment: distance, and horizontal and vertical tilt.
- High, straight slits provide a high resolution and a large aperture.

Disadvantages

- The entrance and exit are extremely close to each other.
- A high risk of internal reflections and multiple dispersion at low grating angles.
- A high risk of stray light.

2.3.2 Ebert–Fastie configuration

Hermann Ebert invented his spectrometer in the 1890s, and his concept was efficiently improved by William Fastie in the 1950s. Ebert avoids Littrow's main problem, i.e., the close vicinity of the entrance and exit, by placing them on two sides of the dispersing element, far enough that their beams do not touch the grating edge. However, a much wider mirror is required, and the entrance and exit show a small tilt angle versus the perpendicular of the front plate of the instrument (this allows the two distances, mirror–grating and mirror–entrance, to be of equal length).

A serious disadvantage of the original Ebert concept is the possible direct reflection of light from the entrance to exit without diffraction. Fastie modified the concept at several points: He increased the width of the mirror to allow a baffle (light trap) in the center (thus avoiding the direct transfer from entrance to exit). He invented the curved slits located at the radius around the grating's center, which is also the center of the instrument, while leaving the mirror at the same axes. This design provides stigmatic transfer even at high slits. All rays entering at the radius of the Fastie circle and reaching the mirror at the required place will be transferred in a stigmatic fashion. Fastie also found that a shortened mirror–grating distance increases the image quality. A distance ~80% of the focal length was found to produce the best image quality. The resulting Ebert–Fastie setup is modern and still comprises narrow

internal angles and ideal symmetry automatically. The grating has separate but symmetric input and output angles, now defined by ε . There is also the grating inclusion angle δ , which is ε twice, and the symmetric out-of-plane angles ι . This setup validates the general grating equation [Eq. (2.1)], but because the system is symmetric, one can also apply the Littrow equation [Eq. (2.6)] and add a correction factor, which is the cosine of ε , leading to the Ebert equation>

$$m \times \lambda = k \times 2 \sin \alpha \times \cos \varepsilon. \quad (2.8)$$

Instead of α , the working angle Φ is used for further calculations:

$$m \times \lambda = k \times 2 \sin \Phi \quad (2.9)$$

and

$$\Phi = (\alpha + \beta)/2, \quad (2.10)$$

because it is a symmetric system.

Using $\alpha + \beta$ also fulfills the basic rules, but Φ is more general and fits better with the dispersion calculations (discussed later in Section 2.6.4). When calculating the grating angle for a given wavelength, the equation results in the working grating angle. Thus, α is the calculated angle Φ plus ε , whereas $\beta = \Phi - \varepsilon$. This does not produce the same result as omitting $\cos \varepsilon$ from the calculation. Φ is the angle at which the grating is moved away from the position of reflectance (zeroth order).

Equation (2.1) is only required if $\alpha < \varepsilon$, and β might turn negative because it will be on the other side of N . Because of the slim angles, Ebert–Fasties are often used for long focal lengths. Consider the following example: $f = 1500$ mm, with a grating of 140 mm \times 140 mm, giving a theoretical aperture of $f/10.7$, a rather “fast” system at that focal length. There is an aperture cone angle of ~ 10 deg. The slits need to be at least 30 mm apart from the grating’s periphery, resulting in a distance of 220 mm between the entrance and exit. The mirror will be designed with some reserve, as both areas have a 145-mm width. Because a beam stop of 20 mm will be added in the center, the mirror will be 310 mm wide and 140 mm high. The entrance beam will hit the mirror 82.5 mm beside the instrument’s center, which causes the slits to be slightly tilted: 1.05 deg relative to the front plate, as over-stated in Fig. 2.6. That small of an angle does not require a slit to be tilted, but it requires an area detector to sit on a wedged adapter. The median mirror-grating distance following Fastie’s rule is 1215 mm, producing an ε of 6.1 deg or a correction factor ($\cos \varepsilon$) of 0.994. The out-of-plane angle ι on both sides is 3.9 deg, which needs to be factored into the calculation of the mirror curvature but has no impact on the transfer quality. The area-correction equation is applied as follows:

As mentioned previously, only the Littrow setup requires correction in two axes. All other concepts described spread in one plane only (the horizontal, in most cases). Of course, if one folds a beam to save footprint,

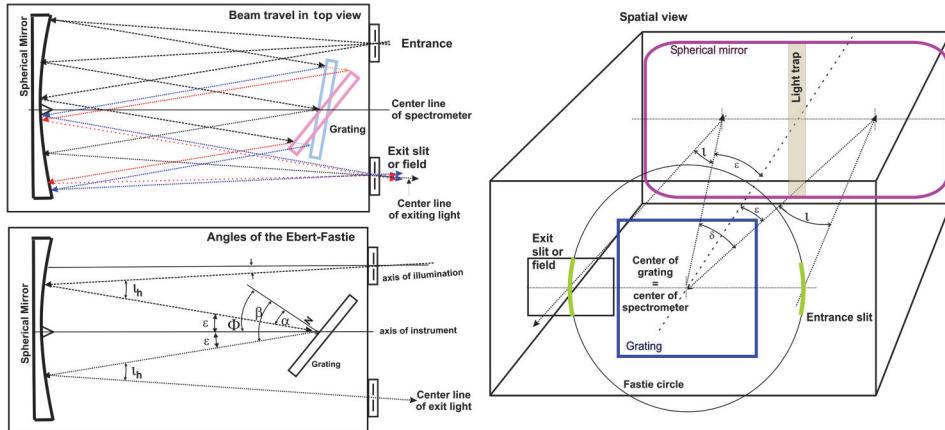


Figure 2.6 Ebert–Fastie setup seen from top and spatial views.

the vertical axis might become relevant again. Thus, there is the Ebert–Fastie area-correction equation:

$$W_{iG} = W_{iM} \times \cos \alpha \times \cos \nu_{h1}, \quad (2.11)$$

where W_{iG} is the illuminated grating width, W_{iM} is the illuminated mirror width, α is the grating's illumination angle relative to N , ν_{h1} is the horizontal out-of-plane angle in the entrance beam, and is the output analogy:

$$W_{M-out} = W_{iG} * \cos \beta * \cos \nu_{h2}, \quad (2.12)$$

where W_{M-out} is the illuminated output mirror width, β is the dispersion angle, and ν_{h2} is the horizontal out-of-plane angle in the exit beam.

The ray traveling inside a spectrometer, among other parameters, is distorted by the height of the slits. An Ebert–Fastie system is symmetric by design and thus ideal for the application of curved slits, which compensate for the differences in ray travel. Like the Littrow, the Ebert–Fastie has only three degrees of freedom in adjustment and is very well suited for attachments (a double-pass option or up to two entrances and three exits are available). It also makes for superb multiple-stage systems. All types of gratings and prisms can be mounted, especially Echelles, which call for $\epsilon < 10$ deg. (Section 2.6.1 discusses the curved slit further.)

Advantages

- Very good stigmatic transfer, caused by symmetric design;
- High degree of flexibility;
- Works with all kinds of dispersers;
- Large curved slits offer high resolution combined with high light flux; and
- Only one mirror.

Disadvantages

- Risk of internal reflection and multiple dispersions at grating angles near ϵ , and
- Problems can appear in image reproduction with detectors >6 mm in height if no extra correction is applied (see Section 2.4).

2.3.2.1 Origin of astigmatism

“Stigmatism” is the expression for the fidelity of a point-to-point reproduction of an object at a monitoring surface. The expression “astigmatism,” referring to stigmatic errors, is used to define the distortions in a photograph, a microscope, or any other optical reproduction. The stigmatic errors typically increase with the distance from the center of the illuminated surface and also with the distance from the entrance’s center spot. Single-axis (lens) systems are based on changes to the refractive index. In reflecting systems, the changes are due to the internal angles of the system. In optical systems with different angles in the vertical and the horizontal plane (which is true in spectrometers except Littrows), the focal lengths are different for the two planes. The tangential focus is reached at a shorter distance than the sagittal focus if the spread is horizontal, i.e., if the entrance, grating, and exit are side-by side. If the system is used as a monochromator, only the tangential focus is important so long as all rays reach the surface of the single-point detector. If used as a spectrograph, both foci should ideally appear at the same place. The wider the internal angles of the spectrometer are, the stronger the difference in focal lengths of the two planes, and the smaller the acceptably illuminated field. (Techniques for correcting astigmatism are described in Section 4.1.6.)

2.3.3 Czerny–Turner configuration

Czerny and Turner presented their concept in the 1930s; it is based on the Ebert design, but the mirror is separated into a collimator and a focusing element, which is also called the camera mirror. Figure 2.7 illustrates this setup.

The Czerny–Turner concept offers much more variety for construction. The direct reflection of input to output is prohibited. Both mirrors can be made of different sizes, which has advantages when illuminating wide areas. In principle, both mirrors can even have different radii, but that causes more harm than good, mainly because of the change in focus. Furthermore, they can be made toroidal to allow stigmatic transfer with straight slits. The beam configuration can be designed such that all entrances and exits are plane parallel to the housing. The grating might be off the symmetry axis of the instrument; however, Czerny–Turner has nine degrees of adjustment freedom, which can create some problems. A clear advantage is that the freedom of building the instrument is much wider than an Ebert–Fastie with the same focal length and f -number, resulting in wider ϵ and δ angles. This design

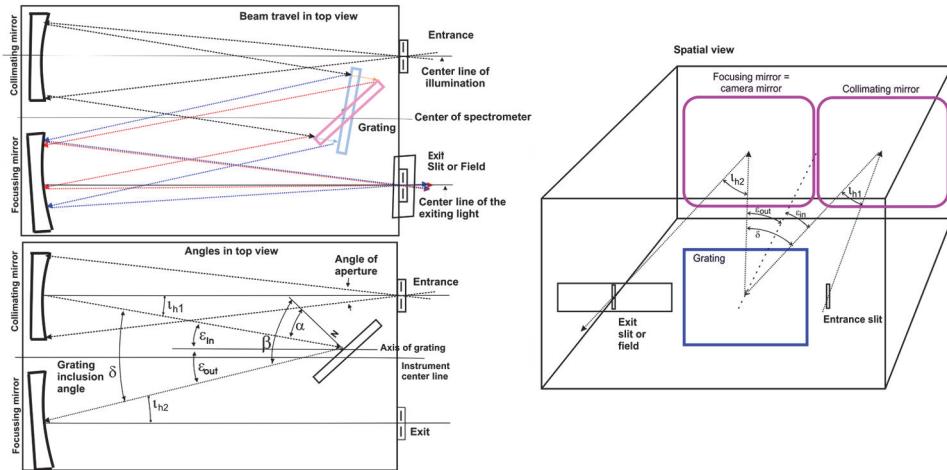


Figure 2.7 The Czerny–Turner spectrometer and its angles.

minimizes internal reflections and multiple dispersions, and eventually produces less stray light.

At the same time, as the angles α , β , and ϵ and the inclusion angle δ increase, the grating function and aperture are altered, which leads to decreased imaging transfer quality or more complicated correction. Furthermore, at $\epsilon > 10$ deg, Echelle gratings will not work properly anymore. If the monochromator application requires high slits ($>f/20$), they should be curved, as described in Section 2.6.1. This, in turn, requires a symmetric design that suffers as the angles increase. When performing imaging experiments in spectrograph mode, it is advised to use a system with imaging correction, but even then, cushion or barrel aberration will result from different traveling distances in wide instruments. A symmetric design and narrow angles will help in any case to minimize the required corrections for stigmatic transfer.

The effective area is calculated by Eqs. (2.10) and (2.11), analogous to the Ebert–Fastie setup.

2.4 Impacts and Distortions to Spectrometers

By applying Eqs. (2.1) and (2.8), one finds that the working angle ϕ describes the angle halfway between α and β . This parameter is important in the calculation of dispersion, as will be discussed in Section 2.6.4.

The equation can be described as follows:

$$m \times \lambda = k \times 2 \sin \phi. \quad (2.13)$$

The basic design can influence the internal angles. Consider the influence of the internal angles on the output wavelength of an Ebert–Fastie versus a Czerny–Turner. If ϵ_1 and ϵ_2 are equal, the Ebert–Fastie equation applies in both cases [Eqs. (2.8) and (2.9)]. α and β are positioned above and below, respectively, the

calculated angle by ε . If ε_1 and ε_2 differ by more than 5 deg, the calculation will lead to errors. At this point, the basic equation [Eq. (2.1)] applies.

2.4.1 The influence of the internal angles on the wavelength

Compare three slightly different configurations: With 500 nm in the first order, a grating of 1200 mm^{-1} (or $k = 0.833 \mu\text{m}$) turns toward the entrance. Ebert–Fastie and Czerny–Turner have an inclusion angle $\delta = 30 \text{ deg}$, although the latter is analyzed for a symmetric and an asymmetric setup. A Littrow setup, according to Eq. (2.6), would be

$$0.5 \mu\text{m} = 0.833 \mu\text{m} \times 2 \sin \alpha, \text{ or } \sin \alpha = 0.5 \mu\text{m} / (0.833 \mu\text{m} \times 2) = 0.300,$$

resulting in $\alpha(\text{Littrow}) = 17.45 \text{ deg}$.

For an Ebert–Fastie or symmetric Czerny–Turner [according to Eqs. (2.8) and (2.9), respectively] with presets $\varepsilon = 15 \text{ deg}$ and $\lambda = k \times 2 \sin \Phi \times \cos \varepsilon$, the results are

$$0.5 \mu\text{m} = 0.833 \mu\text{m} \times 2 \sin \Phi \times 0.9659, \text{ or}$$

$$\sin \Phi = 0.5 \mu\text{m} / (0.833 \mu\text{m} \times 2 \times 0.9659) = 0.3136,$$

and thus $\Phi(E-F) = 18.28 \text{ deg}$. Because this would be the working angle, α and β will differ from that by the value of ε :

$$\alpha = 18.28 \text{ deg} - 15 \text{ deg} = 3.28 \text{ deg},$$

$$\beta = 18.28 \text{ deg} + 15 \text{ deg} = 33.28 \text{ deg}.$$

Next, there is a countercheck by Eq. (2.1):

$$\lambda = k(\sin \alpha + \sin \beta) = 0.833 \mu\text{m} (0.0572155 + 0.54873) = 504.7 \text{ nm}.$$

Even with slight changes in configuration, a difference of 4.7 nm to 500 nm is the result, and even that might be acceptable.

Now consider an asymmetric Czerny–Turner where $\delta = 30 \text{ deg}$, but $\varepsilon_1 = 10 \text{ deg}$ and $\varepsilon_2 = 20 \text{ deg}$, because the grating sits out of center and the mirrors are of different sizes. Using Eqs. (2.8) and (2.9) leads to the same result as above. Correlating with ε_1 and ε_2 produces $\alpha = 8.28 \text{ deg}$ ($18.28 \text{ deg} - 10 \text{ deg}$) and $\beta = 38.28 \text{ deg}$ ($18.28 \text{ deg} + 20 \text{ deg}$), respectively. Countercalculating those numbers produces $\lambda = 0.833 \mu\text{m} \times (0.1440 + 0.6195) = 636 \text{ nm}$, which is obviously wrong. The reason for this is the working angle, which is different. This example proves that to perform correctly, asymmetric Czerny–Turners, among all other asymmetric models, must be calculated by Eq. (2.1) exclusively. (See Section 4.7.1 for details about the methods for correcting the image at the x - y output field.)

2.5 Other Spectrometers, Including Those for the Vacuum Range

Systems with a curved grating reduce the number of reflections, which minimizes cost, losses, and probably the stray light level. Because the

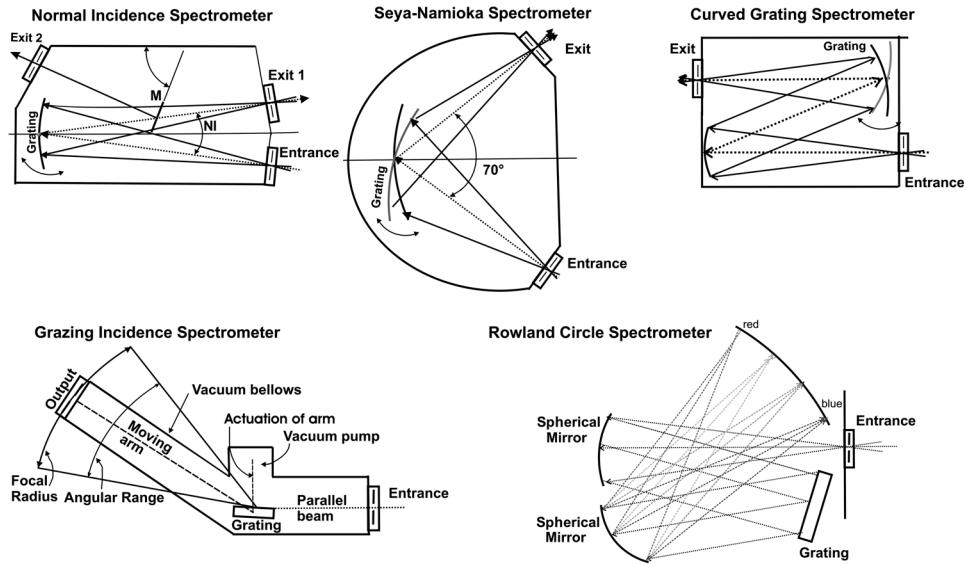


Figure 2.8 More spectrometer concepts.

curvature of the grating modifies the image of the focus in the output field, curved grating systems are mainly used as monochromators. For spectrograph usage, the grating needs a special curvature and can only be used in a narrow angular range. Thus, curved-grating spectrographs are popular for narrowly defined applications, such as production or quality control. In the deep UV, especially below 150 nm, the reflectivity and transmission of all materials regularly used in optical spectroscopy drop quickly. The absorption of air limits applications below 190 nm so much that purging or evacuation is required. Standard quartz materials can only be applied down to about 170 nm. Down to ~150 nm, windows and lenses made of sapphire can be used. Above 106 nm, MgF_2 or CaF_2 are useful. Below 106 nm, no technically useful material transmits, and thus spectroscopy needs to be applied in open systems under vacuum and by reflecting components. The wavelength range below 190 nm can be served by nitrogen purging to get rid of the water and oxygen absorption; the same method is advised for the 1100–2800-nm range, as well, due to the water absorption bands. Below 150 nm, the system requires more and more purging with extremely clean gas to provide reasonable transmission, at which point it might be easier and cheaper to apply vacuum pumps for evacuation. Therefore, different spectrometer versions have been developed. They are efficient down to 106 nm, and in some versions down to 30 nm (40 eV). Photon energies higher than that (shorter wavelengths) involve completely different concepts that are beyond the scope of this book.

2.5.1 Curved-grating spectrometer: Wadsworth setup

In a Wadsworth design, a curved grating reduces the number of reflections to two, and the models are often rather small, with a short focal length. The entrance cone is the same as in the other setups described earlier. The grating has a dual function: it disperses the light and refocuses it to the exit slit or a limited field. Because the entrance and exit are opposite each other, rather large periphery can be mounted to relatively small spectrometers. The Wadsworth spectrometer works with two concave reflectors, and thus the angles are limited in variation, and the field is small.

2.5.2 Normal incidence

To cope with declining efficiencies, spectrometers with no mirrors have been designed besides the Czerny–Turner principal. In these cases, the curved grating has a triple function. It is curved to add both mirrors' functions to dispersion. The inclusion angle δ of normal-incidence (NI) instruments is 15 deg or less. Up to this value, the rules of specular reflection are valid. Thus, there will be negligible change in the state of polarization, high area efficiency, and good image transfer. However, spectrometers with short focal lengths offer only a small space between the entrance and exit for adding periphery. Optionally, several NI systems offer a flat deviation mirror (M) between the grating and exit; the unit has one entrance and two exits. The image transfer is not overwhelming, but in most cases it is good enough for line detectors of up to a 12.5-mm width and 2.5-mm height. NI spectrometers work from 30 nm to NIR.

2.5.3 Seya–Namioka

As an alternative, the Seya–Namioka setup can be applied. The entrance and exit are under 70 deg, offering plenty of space to attach accessories or connect to vacuum systems. With the increased δ , the dispersion also increases at the expense of the projected grating surface: at direct reflection, it is down to 66% of the grating's surface, and it will decrease with the grating movement. Furthermore, the grating's curvature needs to be designed for the instrument's angles. Due to the large difference between the entrance and exit angles, shadowing of the grating structure can become a problem because the groove modulation will be shallow. Polarization effects need to be considered at the high angles given. Finally, the transfer function will be poor (except the grating will be corrected), providing a good image for a very limited range of working angles that, in turn, might require a special slit shape. Regardless, imaging applications will be very poor. Seya–Namioka spectrometers are also available down to \sim 30 nm.

2.5.4 Grazing incidence

In the high-energy range of 1–30 nm (1200–400 eV, soft x ray), the concept of grazing incidence is introduced. The grating will be illuminated with parallel

light under angle $\alpha = 82\text{--}88$ deg. The entrance aperture does not have to focus but rather limit the size of the collimated beam. The detector will move on a radius to collect the required set of photon energy. If focusing is required, a hollow metallic mirror will be mounted at the radius and projected at the detector.

2.5.5 Rowland circle spectrometer and Paschen–Runge mount

A spectrometer with two spherical mirrors and a plane grating refocuses the separated wavelengths at the sector of a circle known as the Rowland circle. All wavelengths directed at the circle (and within a certain angular coverage) are well refocused. The method is realized by two different detector arrangements: The detector can travel on a rail and be moved to the wavelength of interest, or, more often, a series of detectors is fixed at the circle to receive a pre-defined wavelength; this was the standard in emission spectroscopy for many years, up to the 1980s. If the grating is concave, the second mirror is obsolete, and the performance is similar but less flexible. That version is named the Paschen–Runge mount. Vacuum-compatible or purged systems have been used routinely for the range of 170–900 nm.

2.6 Other Parameters and Design Features

2.6.1 Straight slits versus curved slits

Spectrometers are optical instruments: they like symmetry. Imagine a straight entrance slit with a 25-mm height; it can be easily understood that rays entering at the upper or lower end of the slit have a longer distance to the center of the grating compared to the rays entering the center of the slit. A calculation at two focal lengths, combining two distances of the slits from the center of the instrument, can help. The focus displacement results from the trigonometric rule, which includes r_1 , the radius of the Fastie circle; r_2 , the theoretical radius at the end of the straight slit; and d , the distance examined from the slit center. The displacement occurs from the radius and distance to the slit center, not from the focal length, and thus longer-focal-length instruments come with a wider radius to enable larger gratings. The instruments compared (A and B) are as follows:

- A. focal length = 1 m ($f/10$); $r_1 = 110$ mm; $r_2^2 = (r_1^2 + d^2)^{1/2} = ((12100 + 156.25) \text{ mm}^2)^{1/2} = 110.708$ mm.
- B. focal length = 0.3 m ($f/5$); $r_1 = 80$ mm; $r_2^2 = (r_1^2 + d^2)^{1/2} = ((6400 + 156.25) \text{ mm}^2)^{1/2} = 80.971$ mm.

The displacement will cause the focus reproduction to deteriorate. Its shape depends on other parameters and angles; in the worst case it will just add to the slit width (as shown in Fig. 2.22). The error increases from the center to the periphery of the slit, but it is constant at a certain distance and thus causes more difficulties with narrow slits or fine images. At shorter slits, the impact is greatly reduced. For examples A and B, at ± 2.5 mm, A = 2.3- μm displacement, and

$B = 3.1 \mu\text{m}$. Because that size is below the Rayleigh diffraction limit, it will not cause any problems, which is the main reason for limiting the slit height in spectrometers with straight slits.

There are other solutions to this problem. Apart from installing curved slits, which require an axial-symmetric spectrometer design, the slit height used can be reduced until the resolution is acceptable, at the expense of light flux. For imaging spectrometers, optical correction methods are available. (See also Section 2.6.7.5.1 on coma and Section 4.1.6 on output fidelity and correction methods.)

2.6.2 Aperture and light flux (luminosity)

A general equation is used for luminosity or light flux in optical systems:

$$L = A^2 \times T \times \Omega, \quad (2.14)$$

where L is the luminosity (an arbitrary number), A is the illuminated area of entrance and exit of the system, T is the transmission through the instrument at the wavelength under consideration, and Ω is the aperture of the beam. It consists of the active internal area A divided by the square of the focal length, thus taking into account that the light density decreases with the square function of the focal length.

Ω represents the normalized cone angle through the spectrometer; it reflects the fact that the light density (radiane) at a given area drops by the square of distance:

$$\Omega = A_g/f^2, \quad (2.15)$$

where A_g is the illuminated grating area, and f is the focal length that produces the light flux through a spectrometer, described by the arbitrary number L , which allows for the comparison of different instruments for their applicability to a defined experiment.

These parameters are discussed further in Chapter 6. Ω plays a key role in illumination (Section 6.3) and in related applications.

2.6.2.1 Real aperture or f-number?

As with a photographic camera, the relation between the active surface and focal length, the aperture, is a measure for the capability of a spectrometer to collect light and guide it to the output (often called the “light-collecting power”). The shorter the focal length and the larger the optical components involved, the more light can be transferred. Also, analogous to other optical systems, the quality of reproduction drops with increasing Ω , which also leads to decreasing spectroscopic performance due to the increasing internal angles of the system. Also note that the chance of getting unwanted light in the output increases.

The aperture is defined by $n = f/Wi$. It is also called the “f-number” or “f-ratio.” The light flux in a system of $f/4$ is not twice the light flux of $f/8$: it is $(f/4)^2/(f/8)^2$. A spectrometer with a focal length of $f = 300$ mm and the

capability to illuminate a grating 100 mm wide has $n = 300 \text{ mm}/100 \text{ mm}$, or $n = 3$. Unfortunately, by no means it is assured that the grating provides $n = 3$ in real use. Most gratings are rectangular; one with side lengths of 70 mm \times 70 mm has a diagonal of roughly 100 mm, but calculations would include nonexistent segments (see Section 2.6.2.3). Some manufacturers specify the spectrometer as if the areas existed anyway. In other words, if the grating is 100 mm \times 100 mm, the diagonal is 140 mm, and some existing segments will remain unilluminated. However, is that really a problem? For both systems, the f -number in the datasheet might be the same! It is already known that area losses rise as the grating turns.

Consider a Czerny–Turner with $f = 300 \text{ mm}$ and inclusion angle $\delta = 30 \text{ deg}$. The grating is 70 mm in width and height, making a diagonal of 100 mm. The mirrors provide some reserve and have a 110-mm diameter. It is obvious that the projected height remains stable as the grating turns, but the width does not. Based on δ , the mirrors are tilted versus the axis by 7.5 deg, producing a loss of $\cos 7.5 \text{ deg}$, or <1%. The grating “sees” both mirrors at an angle of 15 deg, which creates another loss of 3.5% in both directions, even at zero order. Therefore, it is a good idea to use a wider grating. If it rotates towards the entrance, it reaches full projection at $\alpha = 15 \text{ deg}$. Above that angle, the illuminated area decreases again. At the illumination angle $\alpha = 45 \text{ deg}$, it will suffer from 30% loss, at which point only two-thirds of the optimum aperture ($n = 4.5$) are used. If the dispersed light uses the output mirror fully, then the real aperture has been found. If not, the real aperture will be even weaker. Users are advised to not blindly believe the aperture numbers provided by datasheets; rather, they should collect the component sizes and angles, and perform some calculations for planned experiments. The use of Ω will be very helpful—if the Ω of the illumination and the spectrometer are identical, the losses are minimal, and the spectrometric performance is the best.

2.6.2.2 Examples of the influence of the internal angles on the light flux

Consider the following comparison of similar spectrometers with different distances between the collimating and focusing mirror, as well as between the entrance and exit:

- **Center distance = 250 mm**, mirrors and grating are 100 mm \times 100 mm, and $f = 500 \text{ mm}$. The distance between the grating’s center to a virtual point between the mirrors is assumed to be 400 mm; with a symmetric ε and a test wavelength of 500 nm, the grating rotates towards the entrance. The theoretical aperture is $n = 5$. For the 25-cm-wide unit, $\alpha = 17 \text{ deg}$, and $\iota_h = 13 \text{ deg}$. The projected width of the collimator drops from 100 mm to 97.43 mm. At a 25-cm distance between the beams, $\beta = 44 \text{ deg}$. Consequently, the projected width at the focusing mirror is 70.1 mm, leading to a working aperture of $n = 7.14$.
- **Center distance = 300 mm**, all other parameters match the previous example, including the theoretical aperture $n = 5$. However, ι_h is now

19 deg, shrinking the effective width of the collimator to 94.55 mm. Also, $\beta = 57.8$ deg, and the projected width at the focusing mirror drops to 53.3 mm. Because the grating itself also has a smaller projection, the total used width becomes 50.4% of the basic size, i.e., a working aperture of $n = 9.9$. In both cases, the used height remains at 100 mm.

In short, there are good reasons to keep a spectrometer as slim as reasonable and to add some experimental reserve to the width of the grating.

2.6.2.3 Calculating the f -number versus Ω for light flux/luminosity

Even after exact calculation of the f -number n , only half the work is done because the efficiencies of the components at the considered wavelength are not yet provided. Additionally, there is another parameter: the illuminated areas of the entrance and exit.

Unfortunately, the f -number is not clearly defined. Knowing the number does not directly lead to the throughput of a spectrometer. On the other hand, the light flux calculation based on Ω does lead to a basis for comparison. The grating area A (the black rectangle in Fig. 2.9), projected from the entrance slit by four solid lines, is calculated. The result is divided by f^2 . Between the collimator and grating, the rays travel in parallel fashion, and both areas are the same. The resulting Ω reflects the fact that the light density (irradiance) decreases with the square of the distance. Another advantage of this procedure is that it is fully compatible with the definition of steradian used for illumination. In comparison, f -number calculations only regard two linear dimensions: the focal length and one dimension of the illuminated object (mirror, grating).

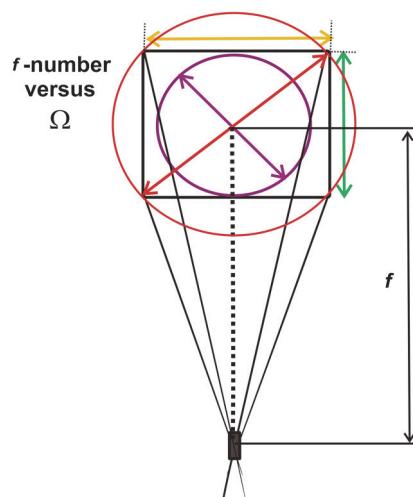


Figure 2.9 Illustration of the f -number in comparison with Ω .

The best fit is reached by applying the grating width (marked by the yellow dart in Fig. 2.9), which allows for observation of the change of size as the grating turns, which unfortunately is not often applied in datasheets. To compensate for the losses, the actual size of a grating (black frame) should be wider than the height (green dart). Sometimes the published *f*-number is based on the inner diameter (purple circle), leading to the same result as if the shorter side were used. This scenario presents an *f*-number worse than reality. To present better numbers, some spectrometer vendors define the *f*-number based on the grating diagonal (outer circle, marked red), but this number promises areas of approximately one-third of the theoretical area, which are, in reality, not present. The majority of datasheets do not document which method is used to calculate the *f*-number. To find the correct number, several calculations are required, and thus it is better to use Ω from the beginning.

Therefore, an alternative method is suggested: the calculation of luminosity, or light flux, based on the general equation:

$$L = A^2 \times T \times \Omega,$$

Based on Eq. (2.14), an equation was developed for spectrometers that contains the relevant parameters to estimate the light flux with monochromators and spectrographs. One version allows for prediction with a suitable spectrometer without accounting for the illumination angle. The other version is based on a known cone angle Ω for the light at the entrance. If the irradiance at the entrance slit is known and put into calculations, the quantitative component of the previous version is added. The normalized light-transfer ratio Ω is

$$\Omega = A^2 / f^2,$$

where Ω is the normalized light transfer ratio, A^2 is the illuminated or emitting area, and f^2 is the square of the focal length or distance.

The luminosity at the spectrometer output (Fig. 2.10) is found by the following equation:

$$L_s = T \times A_s \times \Omega \times (h_D \times B), \quad (2.16)$$

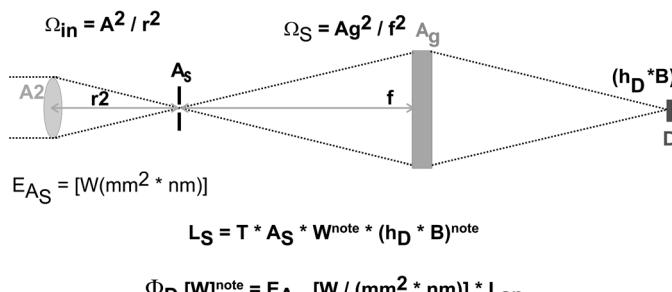


Figure 2.10 The luminosity at the spectrometer output.

where

- T is the transmission of the spectrometer at the wavelength of interest (typically 0.1–0.7). A double spectrometer, comprising identical components, is $T/2$ compared to the single-stage version.
- A_s is the illuminated area of the entrance slit. It is assumed that the irradiance fills the slit completely and uniformly. The dimension is $E_{As} = \text{mW}/(\text{mm}^2 \times \text{nm})$. According to Eq. (2.15), A^2 is the area, assuming that the entrance and exit are of the same size, which is true for monochromators but often not for spectrographs.
- Ω has two meanings. Given as an absolute value, Ω defines the capability of the instrument to collect and transport light, with the radiation traveling in a half-sphere distribution from the slit into the system. However, realistically, the light enters under a defined cone angle, shown in Fig. 2.10 as $A^2/r^2 = \Omega_{in}$. The resulting Ω is the ratio of Ω_S/Ω_{in} , with the result limited to 1 at maximum. In other words, so long as Ω_S is larger than Ω_{in} , all light entering will be transported at the expense of resolution. As soon as Ω_{in} becomes larger than Ω_S , the spectrometer is overilluminated, light will be lost, and even worse, it will create stray light. Hence, it is very wise to adapt Ω_{in} closely to Ω_S , trying to reach a total $\Omega = 1$.
- h_D is the illuminated height in the exit, which can be a slit or a read-out element (pixel or superpixel). It is clear that the maximum height of the exit cannot exceed the height of the entrance. The parameter also includes the radiation that is dispersed at this point. Consequently, the x axis can be defined in millimeters or nanometers, as well.
- B is the spectral bandwidth (nm) in the output slit or at one bandwidth element. The parameter includes the radiation that is dispersed after the grating. The bandwidth results from the dispersion (RD, see Section 2.8.1). If at the time of estimation the bandwidth is not yet known, the geometrical width can be taken. Consequently, the x axis can be defined in millimeters or nanometers, as well.

Note that all calculations negate any kind of aberrations and disturbance—only an ideal system is considered.

The following dimensions are used to calculate a spectrometer regardless of illumination:

$$L_s = T \times A_s [\text{mm}^2] \times \Omega [\text{mm}^2/\text{mm}^2] \times (h_D [\text{mm}] \times B [\text{nm}]) = [\text{mm}^2 \times \text{mm} \times \text{nm}],$$

resulting in an arbitrary output parameter that includes height and bandwidth. If, besides the spectrometer, the illumination cone is also known, the two cones (Ω/Ω) are put in relation. If, in addition the irradiance at the entrance is taken into account, we find:

$$\begin{aligned} L_s &= P [\text{W}/(\text{mm}^2 \times \text{nm})] \times T \times A_s [\text{mm}^2] \times \Omega [\text{mm}^2/\text{mm}^2] / \Omega [\text{mm}^2/\text{mm}^2] \\ &\quad \times (h_D [\text{mm}] \times B [\text{nm}]) = [\text{W}]. \end{aligned}$$

All dimensions except for the optical power are cancelling out, resulting in

$$\Phi_D = E_{As} [\text{W}/(\text{mm}^2 \times \text{nm})] \times L_s = [\text{W}].$$

2.6.2.3.1 Monochromator examples

Two monochromators are compared to find the better one for a defined experiment. In both cases, the original cone of illumination is assumed to be much wider than the acceptance cone of the spectrometer; thus, the half-sphere principle applies.

- **Mono 1** has 110-mm-diameter mirrors, a grating of 70 mm × 70 mm, $f = 500$ mm ($n = 7.14$, $\Omega = 0.0196$), and the internal inclusion angle $\delta = 20$ deg. The slits can be opened up to 3-mm width × 10-mm height.
- **Mono 2** has 110-mm-diameter mirrors, a grating of 70 mm × 70 mm, $f = 1000$ mm ($n = 9.1$, $\Omega = 0.0049$), and $\delta = 10$ deg. The slits can be opened up to 3 mm × 10 mm, as well.

Both share the following parameters in common: a 0.5-nm required bandwidth at the exit (see Section 2.10), a grating of 1200 mm^{-1} , an examined wavelength of 550 nm, and a transmission of 50%. Mono 1 has a grating angle of 19 deg and a dispersion of 1.47 nm/mm, leading to a slit width of 340 μm for a bandwidth of 0.5 nm. The resulting slits are 3.4 mm^2 , and the projected grating surface is $70 \text{ mm} \times 0.9455 \times 70 \text{ mm} = 4633 \text{ mm}^2$. Mono 2 has a grating angle of 19.2 deg and a dispersion of 0.77 nm/mm, leading to a slit width of 650 μm for a bandwidth of 0.5 nm. The slit areas are 6.5 mm^2 , and the grating area is $70 \text{ mm} \times 0.9443 \times 70 \text{ mm} = 4628 \text{ mm}^2$.

The following values were applied in the first example (the parameters are in parentheses):

For Mono 1:

$$\begin{aligned} L1 &= (T) 0.5 \times (A_s) 3.4 \text{ mm}^2 \times (A_g) 4634 \text{ mm}^2/(f^2) 500^2 \\ &\quad \times (h_D) 10 \text{ mm} \times (B) 0.5 \text{ nm} = 0.158. \end{aligned}$$

For Mono 2:

$$L2 = 0.5 \times 6.5 \times 4628/(1000^2) \times 10 \times 0.5 = 0.075.$$

For quantitative calculations, the L values only need to be multiplied by the beam density (irradiance) at the entrance [$\mu\text{W}/(\text{mm}^2 \times \text{nm})$] to receive the radiant power (μW) in the exit. Isn't that an elegant tool? Despite the f -number of Mono 2 being worse by a factor of 2 (which leads to $\Omega = 1/4$), half of the light comes out and is received by the detector. Other advantages include a better contrast ratio, better image transfer, and higher flexibility (experimental reserve). Consequently, when comparing monochromators, the L calculation is far better than trying to determine the better system by f -numbers alone.

If the illumination system is fully adapted to the spectrometer, the Ω will consist of a normalized figure between 0 and 1, resulting from Ω_S/Ω_{in} . Keep in

mind that the light is no longer lost inside the spectrometer, but it will be lost eventually in the illumination system. One kind of loss is a strong focus enlargement that does not allow all light to pass through the entrance slit. Thus, recalculation of L1 and L2 (based on $\Omega = 0.9$) is needed, assuming that almost all entrance light is efficiently handled:

For Mono 1:

$$L1 = (T) 0.5 \times (A_s) 3.4 \text{ mm}^2 \times (A_{gr}) 0.9 \times (f^2) 1 \times (h_D) 10 \text{ mm} \times (B) 0.5 \text{ nm} = 7.65.$$

Due to the normalization of A_{gr} and f^2 , Mono 2 is found thusly:

$$L2 = 0.5 \times 6.5 \times 0.9 \times 10 \times 0.5 = 14.63.$$

If the quantitative calculation includes the beam density (irradiance) at the entrance in [$\mu\text{W}/(\text{mm}^2 \times \text{nm})$], the result is the radiant flux (which is the luminosity) in μW . Based on this comparison, it can be seen that due to the larger slits, the longer spectrometer is the winner, despite the smaller aperture. It delivers almost twice the light power at the exit.

2.6.2.3.2 Spectrograph examples

As can be seen, the actual slit size has a strong impact on the result of L. Spectrograph applications almost always use area detectors with small elements, which makes it important to control how much light reaches the considered element (pixel or superpixel). The selection of components is difficult in many cases because the bandwidth and interval need to be combined. For comparison:

- **Spectro 1** has, as before, a 110-mm mirror diameter, a grating of 70 mm \times 70 mm, $f = 500$ mm ($n = 7.14$), and an inclusion angle $\delta = 20$ deg. The entrance slit can open to 3-mm width and 10-mm height.
- **Spectro 2** also has a 110-mm mirror diameter, a grating of 70 mm \times 70 mm, $f = 1000$ mm ($n = 14.3$), and $\delta = 10$ deg. The entrance slit can be open to 3-mm width and 20-mm height.

Both examples have the following parameters in common: the spectral interval at the detector is 15 nm, the bandwidth per pixel is 0.05 nm, the grating density needs to be chosen, the reference wavelength is 550 nm, the transmission is 50%, and the detector is 25.4 mm \times 25.4 mm, providing 1024 \times 1024 pixels of 25 $\mu\text{m} \times$ 25 μm .

Spectro 1 has a grating of 1800 mm^{-1} set to 29.2 deg, providing a dispersion of 0.89 nm/mm. This arrangement illuminates a detector width of 16.85 mm to recover an interval of 15 nm. The bandwidth of a single pixel is 0.0225 nm, leading to a slit width of 50 μm and a measured bandwidth of 0.05 nm. The total slit area will be $0.05 \times 10 \text{ mm}^2$, whereas the grating area is $70 \text{ mm} \times 0.873 \times 70 \text{ mm} = 4277 \text{ mm}^2$. The calculation of L1 will show the factor of the output power falling on a superpixel.

Spectro 2 has a grating of 1200 mm^{-1} at 19.2 deg; thus, dispersion will be 0.77 nm/mm, leading to a used detector width of 19.48 mm. A single pixel will

be illuminated by 0.0192 nm, resulting in a slit width of 65 μm for the 0.05-nm bandwidth. The slit area (and the superpixel area, as well) will be $0.065 \times 20 \text{ mm}^2$. The grating area is $70 \text{ mm} \times 0.9443 \times 70 \text{ mm} = 4627 \text{ mm}^2$.

The following numbers apply: For Spectro 1,

$$L1 = 0.5 \times 0.5 \times 4277/250000 \times 10 \times 0.05 = 0.00214.$$

Due to the 10-mm limit of slit height, only 40% of the detector height is illuminated. For Spectro 2,

$$L2 = 0.5 \times 1.3 \times 4627 \times 1.3/(10^6) \times 20 \times 0.05 = 0.003.$$

The detector is not fully used; only 80% of the height is illuminated.

In summary, the system with the weaker f -number wins by a factor of 1.4. All results show that the f -number alone is of no use; it must be combined with several other parameters, which is what the L equation renders. It is up to the reader to try the calculation of Spectros 1 and 2 with the help of a normalized Ω .

2.6.3 Dispersion of spectrometers

In the grating discussion [Eq. (2.5)], the dispersion is defined as $\Delta\beta/\Delta\lambda$. The main interest for a spectrometer, however, is how far two wavelengths are separated in the exit, which means $\Delta \text{ nm/mm}$. Two different wavelengths leaving the grating travel under a constant difference in angle. The larger the distance of the exit from the grating is, the further the signals are separated. Equation (2.5) can be extended to

$$f \times (d\beta/d\lambda) = f \times m/(k \times \cos \beta), \quad (2.17)$$

which includes the focal length. The complete dispersion calculation depends on all of the angles around the grating (Fig. 2.11). With the help of one of the grating equations, the term $(\sin \alpha \pm \sin \beta)$ must be first be found, as well as the internal spectrometer angles ϵ and δ . From all four angles, a new parameter can be found, valid for one wavelength at one order: the grating's working angle Φ , which refers to the grating normal N and is typically halfway

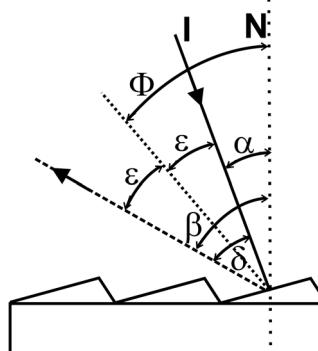


Figure 2.11 The angles at the grating in a real spectrometer responsible for the dispersion.

between α and β . For the calculation of Φ , the following equation can be used:

$$\Phi = \arcsin [\lambda / (2 \times k \times \cos x)], \quad (2.18)$$

where Φ is the working grating angle, defining the median dispersion at a certain wavelength, λ is the wavelength (entering this in μm makes further calculations easier), k is the grating constant (also in μm), and x is the half inclusion angle (in symmetric spectrometers, it is the Ebert angle ε , generally $\delta/2$).

With the Φ found, one can proceed with the calculation of the reciprocal dispersion RD:

$$RD = [\cos(x + \phi) \times k] / (f \times m). \quad (2.19)$$

For a quick overview, the calculation can be simplified to

$$RD = \lambda / (2f \times \tan \phi). \quad (2.20)$$

The function is a reciprocal one because the wavelength is now the nominator; therefore, the value of the ratio nm/mm decreases as the dispersion (the distance between two wavelengths) increases. The RD number decreases with rising mm^{-1} , producing a steeper working angle of the grating, higher spectral order, and longer focal length. Equations (2.19) and (2.20) apply for all spectrometers. The RD found will only fit perfectly in the center of the exit field. The useful field area increases with focal length because of the angle ratio. Precise calculations for array detectors can require corrections toward the outskirts of the field. The expression RLD (reciprocal linear dispersion) is often found in brochures and literature. However, because the dispersion is not linear at any place, the expression RLD is not used in this book; RD appears more reasonable. (The dispersion of prisms is discussed in Section 2.16.6.)

2.6.4 Intensity distribution in the exit

In both monochromators and spectrographs, the intensity of a single band (wavelength interval) in the output is triangular. Assuming a homogenous illumination of the entrance slit with one or several discrete wavelength bands in both x and y orientation, the distribution is shown in Fig. 2.12.

If an optical system is illuminated by a single wavelength and the signal reaches the output spectrally dispersed, the image of the signal in the exit represents the entrance's vertical distribution in a cosine-shaped fashion. In the x axis, where the wavelength is spread, the signal becomes a triangle because the spectral center of the exit represents the center of the singular wavelength. From there, the signal declines in a linear fashion in both directions. When the entrance slit height is filled with uniform light, the exit slit height distribution will be uniform, too, except for the extreme ends, which will decay in a cosine shape. If the entrance slit is illuminated inhomogeneously in both x and y directions, with the intensity concentrated in center and showing a steady decay in both directions, the exit will follow that pattern vertically, but in the horizontal axis it will still be a triangular distribution.

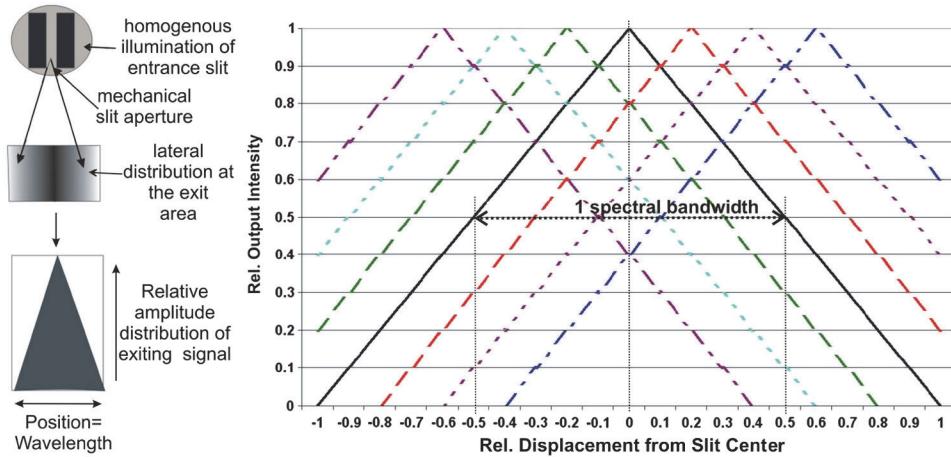


Figure 2.12 Intensity distribution in the exit of a monochromator versus a spectrograph.

The vertical axis at this point is a convolution of triangle and cosine, whereas in the horizontal the wavelength is dispersed. The grating disperses all signals passing through the entrance with the same efficiency. In general, the exit does not know which ray came from which side of the entrance. Therefore, the wavelength set remains in the center of the exit. So long as the detector is not smaller than the exit, no signal will be lost. The output distribution might be important for the application of field detectors, if a monochromator is used as a variable light source, or in radiometric applications. To produce homogeneous output signals, the spectrometer needs homogeneous illumination at the entrance, and it must be stigmatic. If the output is not a slit but a field, it will only carry homogeneous intensity if the focusing mirror is larger than the illuminated area of the collimator; otherwise, vignetting occurs (see Section 4.1.2). This behavior is the reason why instruments with an oversized exit mirror are available.

Due to dispersion, even within the slit width, there is a change of color over the width, too. Figure 2.13 depicts the functions of transfer and output dispersion. In a single-stage monochromator's output slit, the spectral and

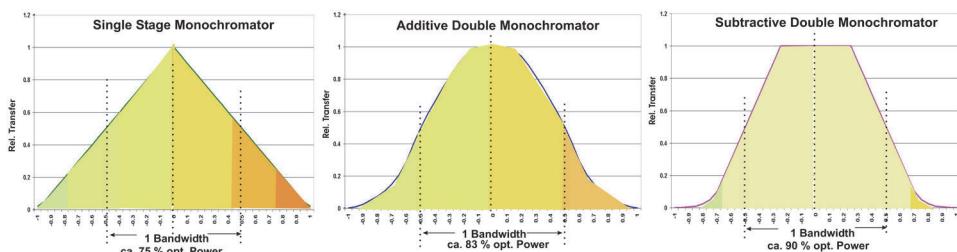


Figure 2.13 Transfer function and distribution of the optical power over the exit slit.

power distribution is shaped like a triangle. The definition of bandwidth is such that one bandwidth covers 50% of the total width, marked by the dotted lines in the figure. The integrated power (if the entrance slit is homogeneously illuminated) is $\sim 75\%$ of the total passing through the output slit. If the “triangular” signal is guided through a second stage with an additive setup (double-pass or double monochromator), the transfer function will be flattened at the top and become steeper. The output bandwidth becomes half the measure of that after the first stage, meaning half the wavelength interval or twice the dispersion. Because of the steeper flanks, the optical power inside the bandwidth increases to $\sim 83\%$. If the second stage is mounted for a subtractive setup, the output becomes trapezoidal due to the inversion of dispersion. The side branches become even steeper compared to the additive setup; $\sim 90\%$ of the total power is now within one bandwidth that will be the same after stage one, but the distribution does not show much internal variation. All energies (wavelengths) from stage one are there, but they are now mixed homogeneously. Only in the outskirts can a small part of different wavelengths appear, which is due to the different pathlength of the rays inside stage two, and lead to negligibly inhomogeneous side wings.

2.6.5 Spectral resolution

Unlike other parameters, “resolution” is not given by a physical law; it is a definition. We can differentiate between the experimental resolution R and a grating’s resolving power R_p . Resolution is determined by the ratio of the center wavelengths λ of a pair of signals, divided by the difference $\Delta\lambda$ between the two signals, which can be safely determined. It is possible to determine resolution if two signals of equal height and very narrow bandwidth are separated by a signal drop of 10% below maximum (see Fig. 2.14).

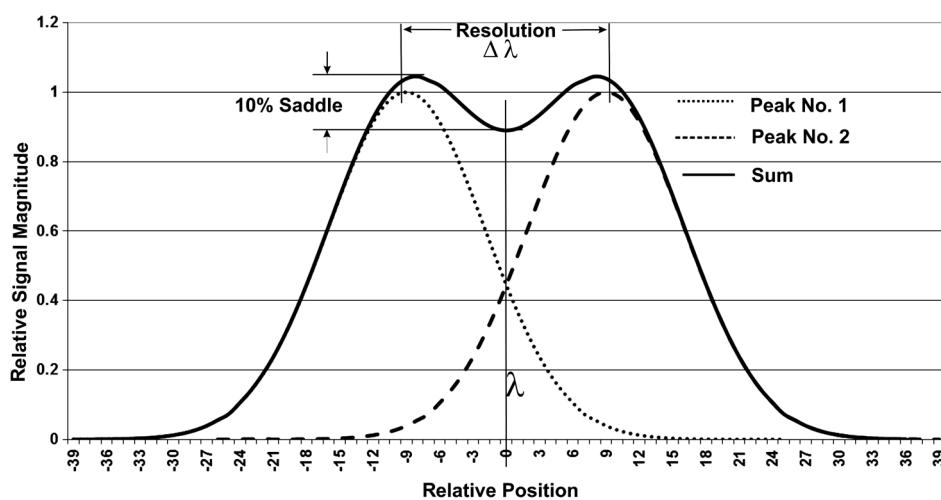


Figure 2.14 Resolution by the 10% criterion.

The graph illustrates the definition of resolution by the 10% rule. The general equation to describe the resolution is

$$R = \lambda/\Delta\lambda. \quad (2.21)$$

The two curves “Peak No. 1” and “Peak No. 2” are of equal amplitude and width, and the saddle between them is at 90% of their maximum. The solid line is the resulting curve. Because both peaks contribute to the signal of the other, the result exceeds the maximum of the single peaks. Note that the two reproduced maxima appear shifted towards the center.

A grating that fulfills Eq. (2.21) has the required resolving power. Theoretically, the resolving power of a grating can be calculated by

$$R_p = m \times W/k, \quad (2.22)$$

where R_p is the numeric theoretical resolving power, m is the spectral order, W is the ruled width, k is the grating constant, and W/k is the total number of lines (W/k is also defined as n).

For example, a grating is 60 mm wide and has 2400 mm^{-1} , $R_p = 144,000$ in the first order. If the actual wavelength is 300 nm, the theoretical spectral resolution is $\Delta\lambda = \lambda/R$, or $300 \text{ nm}/144,000 = 0.00208 \text{ nm}$. In the real world, the achievable performance very much depends on the manufactured quality of the grating and the spectrometer. Regardless, the final resolution R of any spectroscopic setup cannot exceed the calculated R_p .

Now apply the term resolution to λ of the angular dispersion example (Section 2.6.3). There is an angular difference of 0.00215 deg and a main angle of 22.074 deg. The ratio requires a resolving power of $R_p = 10,267$. If that is reversed, it appears that either a smaller grating of 2400 mm^{-1} or a grating with a smaller line density (200 mm^{-1} could theoretically be sufficient) could be used. (This result will return in Section 2.6.6.5, which discusses what R and R_p mean in spectroscopic instruments.)

The term “resolution of a spectrometer” describes the capability to split nearby spectral information without ambiguity. In the earlier discussion of grating resolution, that was assumed to be the case if a saddle of 10% is between two maxima. That is fine for a grating, but for a complete spectrometer to be used for many applications, this scenario might not be sufficient. Unfortunately, some spectrometer manufacturers apply the rule anyway. Looking into several applications and simulating some situations, one can see that the 10% saddle is just not enough. More application-oriented manufacturers apply a saddle of 50% to define resolution, which is a good number. Figures 2.15 and 2.16 show two simulations for the 50% rule. Another way to specify the resolution of a spectrometer is done with a single peak that is much narrower than the spectrometer’s bandwidth, wherein the represented width at half maximum defines the resolution. This situation is called full width at half maximum (FWHM).

Both experiments produce the same result. Resolution is a parameter that is rather easily measured but impossible to be firmly calculated. Estimations

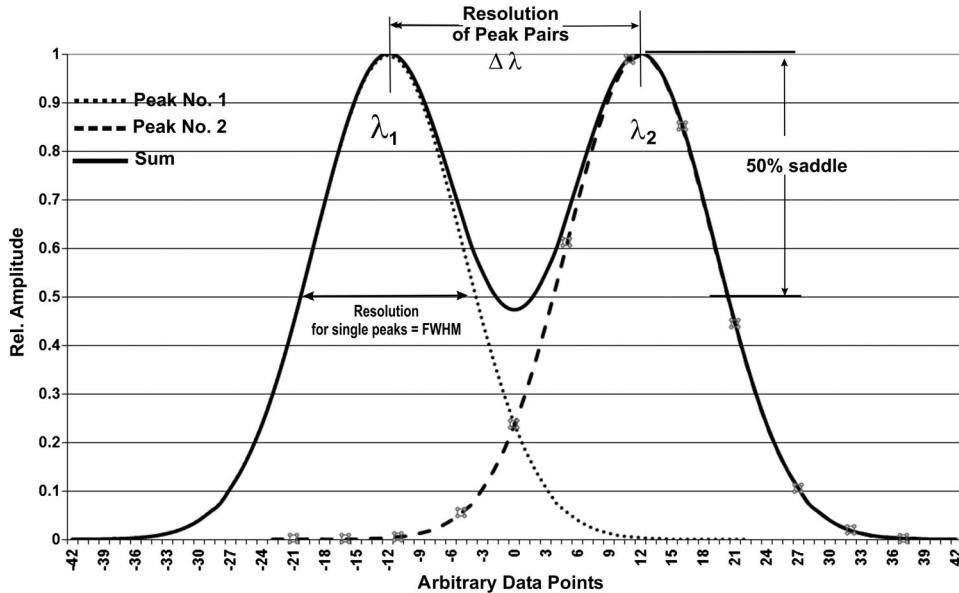


Figure 2.15 Spectral resolution by the 50% criterion.

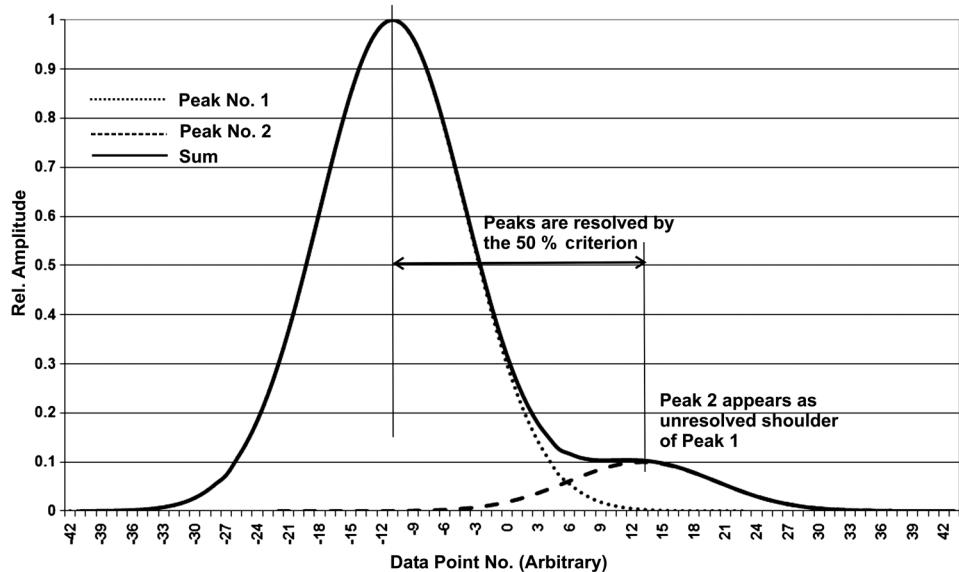


Figure 2.16 Spectral resolution by the 50% rule, and peaks of different amplitude.

only are possible on paper. Resolution improves with the grating's line density, the spectral order, the illuminated grating width, the working angle, increasing focal length, and better surface quality and shape of the beam components. The smaller the entrance aperture is, the better the angular

definition of the entering light, leading to better resolution. The finer the exit slit or pixel is, the finer the spectral interval cut from the parallel interval. Resolution becomes worse with increasing aperture, which is due to the wider internal angles. The rays are less well defined and have more variation, creating competition between wide gratings allowing better resolution versus wide gratings working with wider angles. The working angle presents a similar dilemma: the higher the angle is, the better the dispersion and therefore the resolution, but the difference in traveling distance of the rays at the grating also increases, thus decreasing resolution. The higher the entrance slit is, the stronger the probability for astigmatism in the system, which is why some manufacturers specify resolution with very small slit heights. Therefore, it is important to note the slit height used to specify the resolution when comparing spectrometers. (See Fig. 2.22 in Section 2.6.7.5.1 for more on stigmatic transfer.)

A spectrometer designed for an uncompromised resolution might not be the best choice for an imaging application, and vice versa. A large role is played by the components used, such as the surface quality of mirrors and grating, the homogeneity of the grating's structures, and the quality and parallelism of the slit jaws, not to mention the thermal and mechanical stability of the spectrometer. How do the best components help if the instrument bends or tilts over time due to bad material selection? The experimenter and the environment also influence the resolution. Selecting optimal filters and suitable light sources and detectors will improve resolution. Vibration, dust, smoke, and other environmental drawbacks should be avoided.

Figure 2.15 displays two peaks of equal amplitude and Gaussian shape. Their distance is sufficient for a saddle of 50% in between. According to a widely accepted rule, this defines the peaks to be resolved. Note that the peak maxima are reproduced at the correct wavelength and with the right amplitude.

Figure 2.16 shows the same situation as Fig. 2.15, but peak no. 2 is only 10% the amplitude of peak no. 1. If the 10% saddle rule were applied, the two peaks would be much closer, leading to a better resolution numerically, but the small peak would disappear. If the rule were more rigid, i.e., a 75% saddle, the peaks would be farther apart, but the true position of peak no. 2 would be clear, not like a shoulder of the first peak, as it appears here.

2.6.5.1 Does one measure the resolution of the spectrometer or that of the experiment?

From the curves above, it is unclear whether the result represents the resolution of spectrometer or the experiment. If there are lines much smaller than the spectrometer can resolve, the spectrometer defines the limit. For clarification, a second run is proposed with a distinct reduction in bandwidth (hopefully, the reserve is provided). If the spectrum stays the same, the experiment was the limiting parameter.

2.6.5.2 When is a spectral curve completely reproduced?

The shape of spectral emission and absorption patterns generally appears in two approximations, the Gaussian and the Lorentzian. The chemistry and physics that finally lead to the line shape are discussed elsewhere.¹ In molecular spectroscopy, the Gaussian shape is dominant, whereas in atomic spectroscopy, both shapes are represented. Physical effects on the sample—such as thermal, magnetic, electric, Doppler, Stark, and Zeeman effects—will change the distribution. In most cases, the resulting curve is between Gaussian and Lorentzian. Figure 2.17 presents a reproduction of both.

When calculating the probability for finding the true maximum of a Gaussian or a Lorentzian curve in value and position, one must account for the e -functional signal increase. Considering three data points between the two FWHM values, the maximum will be found with a probability of $[1 - (1/e^2)]$, or 63.2%. The data points must be of equal distance and have a bandwidth of $1/\text{[no. of points]}$ of the FWHM distance. This value ensures that no signal is omitted. The probability increases with the asymptotic function by steps of two. At five data points between the FWHM, one gets 86.6% probability; at seven data points, 95%; at nine data points, 98.2%; and at eleven data points between the FWHM values, the probability reaches 99.33% of the true value $[1 - (1/e^5)]$. That value satisfies most applications.

To describe a peak in maximum, position, and shape, the same number of data points is required above and below FWHM values, producing a sum of 22 data points for Gaussian curves. With that amount of data, the peak will be correct by >99%. In Fig. 2.17, the analysis of a Gaussian curve is shown on the left, with nine data points above FWHM, leading to a satisfactory resolution. On the right is the case of a Lorentzian recursion. The signal distribution of Lorentzian peaks is by far stronger focused on the center. The displayed area represents the ± 5 bandwidths in the center of the Gaussian peak. The “measured” bandwidth is set to one-fourth of that in the Gaussian simulation. This value was chosen to achieve nine data channels between the two FWHM points. It turns out that even a reduced bandwidth might not lead to a satisfactory resolution. Hence, the nine-point rule is sufficient for

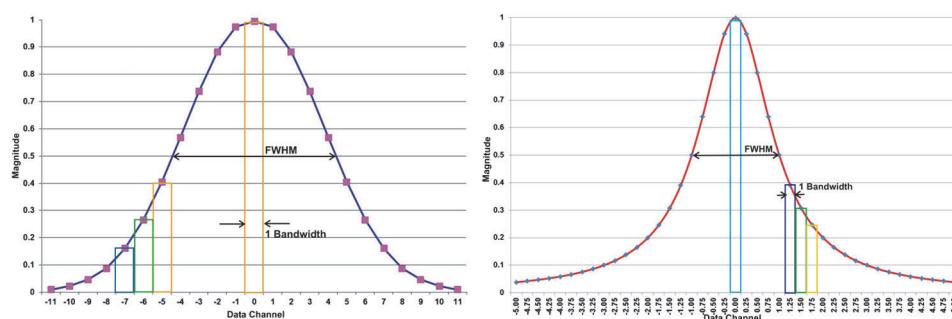


Figure 2.17 Reproduction of a Gaussian (left) and a Lorentzian (right) peak.

Gaussian distributions but might not be enough for Lorentzians, which might call for additional data fitting or more and narrower measured bandwidths.

The vast majority of peaks in practical spectroscopy are Gaussian or close to it. It follows that two Gaussian peaks, no matter how far apart they are, must be separated by at least eleven data points of equal distance and the right bandwidth to capture them with the same accuracy. If the amplitude ratio of the two peaks is $>100/3$, even the eleven points will not be enough because they will not provide a probability of 99.33% for the small peak. The Gaussian curves in Figs. 2.15 and 2.16 have been simulated with >60 data points per peak, which is why the maxima are found perfectly. To reproduce maxima at 98% in position and amplitude, nine points above and six on each side of the FWHM are fine (21 total, which in practice equals the 22 mentioned previously). Figure 2.17 has been created with the same parameters. It assures that the bandwidths do not overlap and that no signal is missed. At four places, the bandwidth is presented as a rectangle. It is easy to see that the bandwidth is too wide for perfect integration at the maximum. From the “bands,” it can be seen how the slopes compensate for the missing and overfilled regions inside the band; the integral is the median value. This is also why the maximum cannot be fully integrated. Even if the band hits the center perfectly, the true maximum is not integrated. It is important that the relations are only valid if the FWHM of the signal measured is at least ten times wider than the applied bandwidth (BW) of the spectrometer.

2.6.5.3 Rayleigh diffraction limit

The resolution improves linearly with a reduction of the slit width. The limit is reached when diffraction at the slit starts to bend the beam, defocusing it. The dimension defining the minimum slit width without extra diffraction is called the Rayleigh diffraction limit. It is calculated by

$$m_s = (\lambda \times f) / (W \times \cos \beta), \quad (2.23)$$

where m_s is the minimum slit width without diffraction at the slit, λ is the actual wavelength, f is the focal length, W is the illuminated grating width, and β is the grating's dispersion angle.

The shorter the wavelength and the sharper the focus in the slit is, the closer the slit jaws can be set. A typical value of m_s in the UV–Vis–NIR is $10 \mu\text{m}$. Longer focal lengths and smaller gratings lead to more parallelism in the beam; thus, the diffraction already starts at wider slits. In spectrograph mode, the entrance slit should be twice the pixel width. If a pixel is wider than $2 \times m_s$, several effects can occur. If the pixels are adjacent to each other in the dimension of the wavelength, binning is required. Otherwise, interference effects between the slit and pixel can occur. These effects can split a single peak into peak pairs, or “ghost pairs,” that might appear close to the original peak if it is very narrow. In any case, the output amplitude will drop below the true value. If the pixels are not adjacent to each other

(or binning is not provided) but their size and/or distance are wider than m_s , the problems are evident, and no solution is possible. Even data-fit routines do not help. The calculated m_s value might be deteriorated if the grating is not ideally illuminated, if the reflecting surfaces are imperfect, if bad slit jaws are used, or if other failures occur.

2.6.5.4 Resolution of a monochromator compared to that of a spectrograph

For a monochromator, a rule of thumb says that the resolution will be 1/20–1/100 of the RD. Consider three examples:

1. A 300-mm monochromator with a grating of 1200 mm^{-1} creates an RD of 2.5 nm/mm at 500 nm. The typical resolution is 0.1 nm, which makes 1/25.
2. A 500-mm monochromator will have an RD of 1.7 nm/mm and a resolution of $\sim 0.04 \text{ nm}$, or 1/40.
3. A 1500-mm monochromator with the same grating has an RD of 0.55 nm/mm and a resolution of $\sim 0.008 \text{ nm}$, or 1/70.

If the signal at the entrance is spectrally much smaller than the spectrometer's band-pass, the signal will not be resolved. The measured curve represents the instrument's resolution. Shifting the wavelength by a two-band-pass distance can already reveal the resolution. One more example: at a slitwidth of $10 \mu\text{m}$ and a bandwidth of 20 pm , one step of 20 pm leads into the saddle, and the next step leads to the next maximum. Therefore, the resolution is 40 pm or 2 bandwidths. The theoretical value for Gaussian distributions is 2.2 bandwidths. If the entrance signal is much wider than the bandwidth of the monochromator, the exit slit will integrate over the wavelengths distributed by the peak while scanning, and the true curve shape will be reproduced if the interval is wide enough.

In spectrograph mode, the minimum slit width is defined by two parameters: the Rayleigh limit and the pixel size of the detector. The slit width must not be smaller than the geometric width of two pixels. Otherwise, the opposite of integration will occur: the convolution of slit and pixel, which can lead to disordered reproduction. A multichannel detector cannot discriminate between two signals illuminating two adjacent pixels because it is not clear where the amplitude originates. To achieve proper discrimination, a minimum must lie between two stronger signals; this arrangement defines the smallest resolution unit to be equivalent to three pixels. The smallest pixel size useful for spectroscopy is about $10 \mu\text{m} \times 10 \mu\text{m}$, making up a minimum geometric resolution of $30 \mu\text{m}$. Applied to the examples above, the minimum resolution is found to be about twice the nanometer value of the monochromator mode, i.e., 0.2 nm for the 300-mm spectrograph.

During peak reconstruction with a monochromator, if the line is finer than the instrument's bandwidth, then the FWHM resolution is ~ 2.2 bandwidths. However, if the line is finer than the instrument's bandwidth with a spectrograph, then the FWHM resolution is at least 3 pixels = 3 bandwidths.

Figures 2.18 and 2.19 illustrate both situations, having been simulated to compare identically built monochromators and spectrographs. A very narrow spectral line—too tiny to be resolved—is recorded. By design, the line does not appear in the center of a bandwidth or a pixel. In monochromator mode, the

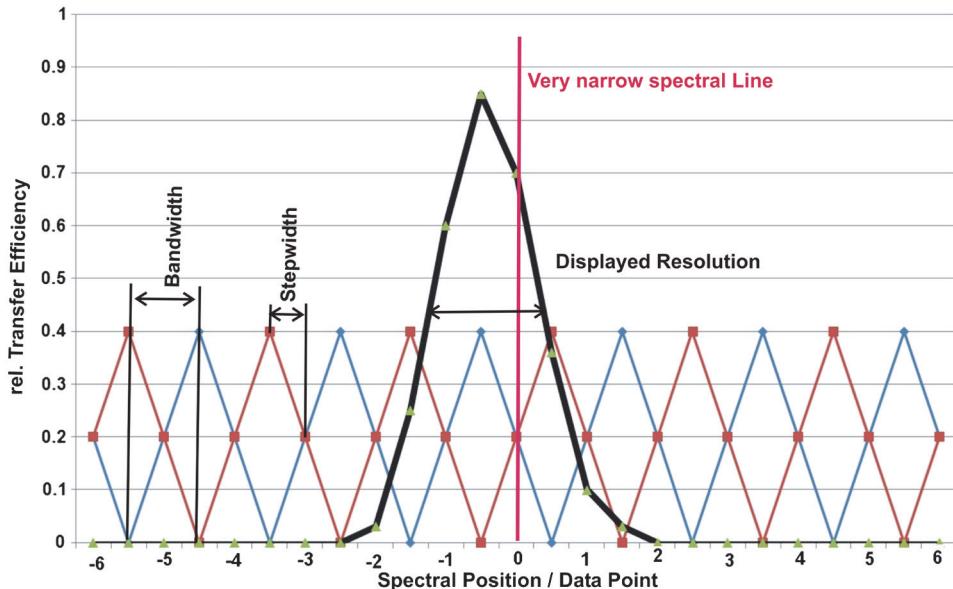


Figure 2.18 Peak reconstruction by a monochromator.

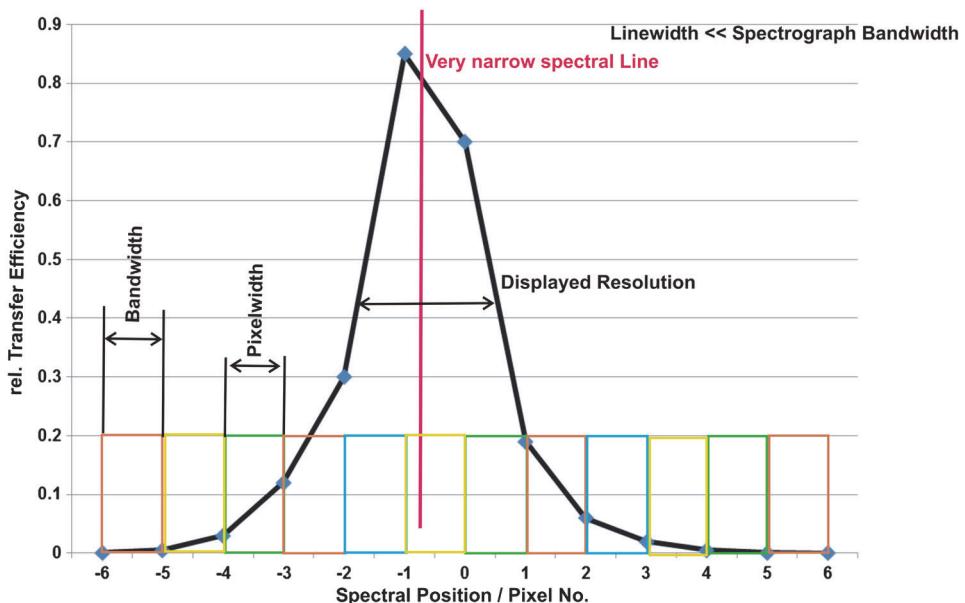


Figure 2.19 Peak reconstruction by a spectrograph.

scan was done in steps of a half bandwidth; in Fig. 2.18, this is marked by the overlapping triangles. As shown in Fig. 2.17, no signal will be lost that way. The reproduction of the line has a resolution of 2.2 bandwidths. If the maximum was in the middle of the center band, the resolution at FWHM would have been 2 bandwidths, which cannot be improved.

In a spectrograph with an array or CCD detector, the pixels form a consistent chain, omitting no arriving signals. In Fig. 2.19, this is shown by the colored rectangles, where each represents one bandwidth. The shape of the “bands” suggests that the reproduction will be different and worse than that of the triangular bands in a slit. The integral within one rectangle rates all signals detected by one pixel equally, whereas in a triangle the center signal has higher weighting than those at the side of the band. Due to the overlap of the triangles, the weighting factor can be between 0.5 and 1.0. Because the area under a triangle and a rectangle is identical, the integral can become the same but only if the bandwidth of the triangle was wider! Furthermore, if a narrow signal lies at or close to the interface between two pixels, the measured amplitude will almost be the same for both. If the line exactly hits the interface, both amplitudes will be identical because a minimum of at least 50% is required to detect the next ascent. Consequently, the minimum resolution possible is the equivalent of 3 pixels = 3 bandwidths in a spectrograph.

Whereas a monochromator typically works with slit widths of 10 μm , and single-point detectors easily reach a linear dynamic range of 10^5 – 10^6 , a CCD or array detector needs to have an area of at least $20 \mu\text{m} \times 20 \mu\text{m}$ to offer a linearity of 10^4 for a single integration. Assuming an identical spectrometer in both modes, using a monochromator will result in a practical advantage in resolution of factor 5 compared to a spectrograph. Comparing their performance under experimental conditions will extend the monochromator’s advantage with better contrast (see Section 2.6.9) and fidelity of signal recursion. The spectrograph, on the other hand, has the clear benefit of recording the same interval with the same resolution within a time advantage of 5–50 fold. Laboratories with instruments equipped with both detection modes can always choose the better method for the actual experiment or correlate both modes to verify the results.

2.6.5.4.1 *Finding the true maximum of a peak*

The definition of the true maximum of a peak produced by a monochromator is found by stepping over the maximum with fine steps, as shown in Fig. 2.17 (Section 2.6.5.2). The results can be verified by repeated scans with a slightly different slit size and step width. The situation is more complicated if the detector is an array. According to Fig. 2.19, the position is fixed during data acquisition. Only when a curve fit unveils that the peak is reproduced fully symmetric does the displayed maximum represent the true maximum position. Because that is seldom the case, the measurement should also be repeated with a slightly different wavelength setting and probably a varied slit width.

Deconvolution software will reduce the true maximum from the stored scans, and at least the data acquisition modes needed to acquire the spectra should be provided by the spectrograph/array vendor. In principle, the same method can be used in spectrograph and monochromator mode.

2.6.5.5 Numerical resolution R_p and R_r , and their wavelength dependence

Even though the example at 500 nm (described in Section 2.6.5.4) is realistic, it is not yet complete. In Section 2.6.5, it was found that the resolving power of a grating is described as

$$R = m \times W/k = m \times W \times [\text{grating}] \text{ mm}^{-1}. \quad (2.24)$$

Equation (2.24) leads to the assumption that the spectral resolution, expressed in nanometers or wavenumbers, will change with energy position. The nm or cm^{-1} resolution is expected to degrade toward lower energies (higher-nm numbers) but improve toward higher energies. Expressed as a ratio, such as $(\lambda)/(\Delta\lambda)$, the result will stay constant; this is called the real resolution R_r . Consider the earlier examples with the three spectrometers in the first spectral order at 500 nm. Now take the theoretical resolving power R_p into account. The ratio R_r/R_p will provide the resolution quality factor Q_r , also called “fidelity of resolution”:

$$Q_r = R_r/R_p. \quad (2.25)$$

For comparison, R_r (300 mm) = 500 nm/0.1 nm = 5000. Because the grating has a width of 70 mm, R_p is 84,000, resulting in $Q_r = 0.06$. That is the difference between the theoretical resolving power and the real resolution in a short instrument. R_p defines only the grating at a virtually infinite focal length, and without other components, it is the limiting resolution.

Back to the examples: R_r (500 mm) = 500 nm/0.04 nm = 12,500 ($Q_r = 0.15$), and R_r (1500 mm) = 500 nm/0.008 nm = 70,000. The latter is already very close to the theoretical value R_p , and Q_r has reached 0.83. In the next step, the R_r values found at the 300-mm focal length are transferred to two more wavelengths. In the UV and NIR it is expected that R_r (300 mm) = 200 nm/5000 = 0.04 nm, and R_r (300 mm) = 1000 nm/5000 = 0.2 nm. The system tends to behave that way, but not fully. Toward the UV, the impact of limiting parameters such as surface diffraction and surface aberration increase and inhibit the calculated 0.04 nm (0.07 will be more realistic). Toward the IR, what is lost toward the UV is gained. By increasing the working angle of the grating, the dispersion grows faster than the sine of the angle. In the end, at 1000 nm, the spectral resolution can be found to be 0.15 nm. This tendency will be similar for all kinds of spectrometers and focal lengths unless higher spectral orders are used.

In short, the quality factor Q_r increases with focal length—rapidly at first, until the focal length reaches about 1.5 m, after which it will slowly approach the value of 1. The spectral resolution is best in the UV and declines constantly toward the IR. One reason for this behavior is that the slit width must be

increased with increasing wavelengths to obey the Rayleigh limit. Instruments with a long focal length in practice achieve more than shown in the examples because long instruments routinely have wider gratings, providing higher R_p , and the surface quality of mirrors and grating in large spectrometers is often better, too. For example, a 1.5-m spectrometer will normally be equipped with a grating 110 mm wide. With a grating of 1200 mm^{-1} , where $R_p = 132,000$, the spectral resolution at 500 nm will be 0.005 nm, or 100,000.

2.6.5.5.1 Practical example of resolution optimization

Consider the wavelength range of 500–600 nm, where an existing 1.5-m spectrograph with a CCD detector calls for optimization. There is a holographic, optimized grating of 2400 mm^{-1} that is 140 mm wide and 120 mm high. The alternative is an Echelle of 316 mm^{-1} with a blaze angle of 63 deg and 220 mm \times 120 mm. The $f = 1500 \text{ mm}$, the detector is a CCD with a 25.4-mm length and 25- μm pixel size, and the calculations are for 550 nm.

The 2400-mm^{-1} grating will work at 41 deg; the real $n = 13.3$; the interval at the detector is 5.1 nm; and the resolution with a 50- μm entrance slit (two pixel widths) is 30 pm. For the Echelle, the working angle in the tenth order is 59.8 deg; $n = 13$; the interval is 2.4 nm; and the resolution will become 14.1 pm, which is better by a factor of 2. The holographic grating is useful for 400–700 nm, and the Echelle between 200–5500 nm. In addition, the transfer efficiency of the Echelle will be at least 25% higher.

2.6.5.5.2 Experimental examples of spectral resolution

The resolution of spectrometers in the UV–Vis range is often tested by acquiring the data of line sources, mainly mercury lamps. A good example is the family of spectral lines at 546.074 nm. Between 546.05–546.1 nm, there are a total of eight lines. The center signal at 546.074 nm contains >80% of the total photons of the group. The discrete lines are only emitted by sources with no or little pressure and thermal broadening. This is eventually the case with low-pressure plasma sources and resonance lamps with extra cooling. Sources with electrodes in the bulb are not useful for resolution tests <5 pm because the small lines will merge; these include high-pressure sources, such as Pen-Rays®. They should only be used for tests >20-pm resolution.

Figure 2.20 illustrates resolution data recorded by several monochromators, each with a different focal length. The darts in the figure refer to the notes in the following description. The reference reading is the bottom spectrum. It was recorded with an Ebert–Fastie double monochromator with $f = 2 \text{ m}$ and a double-pass applied. The effective focal length is 8 m, and the aperture is $f/18$. Each stage has an Echelle grating of 316 mm^{-1} , scanned in tenth order. All three slits have been set to 16 μm , resulting in a dispersion (RD) of 0.02 nm/mm, or $0.67 \text{ cm}^{-1}/\text{mm}$. The system provides a resolution of better than 0.2 pm, or 0.001 cm^{-1} , numerically better than 10^6 . The device tested is an externally excited resonance source with external cooling of the

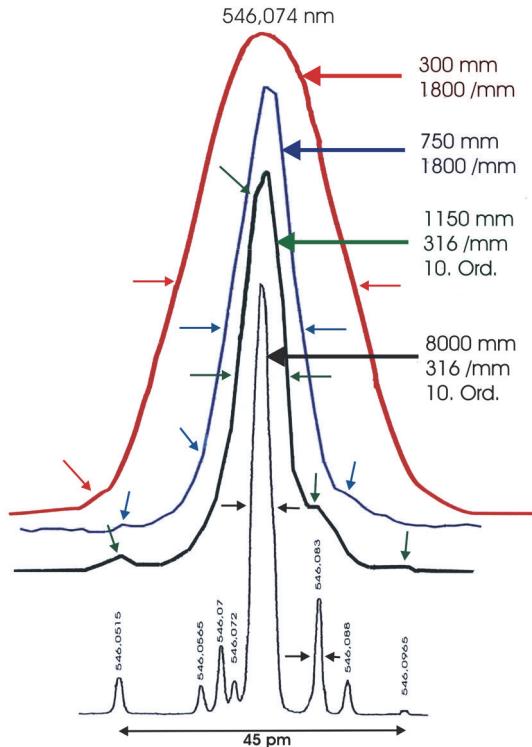


Figure 2.20 Typical resolution data recorded by monochromators of different focal lengths.

body. The tiny peaks are resolved by 0.4 pm. The main peak already shows broadening to 1.2 pm, originated by the source.

The second-lowest curve was taken with an Ebert–Fastie with an 1150-mm focal length, $f/10$, and an Echelle with 316 mm^{-1} in tenth order; the slit width is $10 \mu\text{m}$ at a 40-mm height. The dispersion is 0.13 nm/mm , and the resolution of the instrument is $<5 \text{ pm}$, or $>100,000$. The source is a low-pressure mercury lamp. This specific case reveals two problems: (a) the Hg lines are already thermally and pressure broadened, and they blend together. (b) The monochromator and main line have almost the same bandwidth, which leads to convolution of both functions. The measured FWHM of 8 pm of the central peak incorporates side lines. The leftmost small peak is 22 pm from the center line. Visible but not resolved is the shoulder directly to the left of the maximum. Another shoulder is 9 pm to the right of the maximum, and the final hump on the right has a spectral distance of 22.5 pm to the center.

The second-highest curve was recorded with a 750-mm Czerny–Turner spectrometer with internal imaging correction (toroidal mirror) and $f/11$. The grating is 1800 mm^{-1} , resulting in a RD of 0.63 nm/mm ; it provides a resolution of 20 pm with slits $10 \mu\text{m}$ wide and 10 mm high (30,000). The test object is again a low-pressure mercury lamp. The central peak represents the

resolution of the monochromator; the side lines are indicated but not resolved. Only the small peak 22 pm left of the center is clear.

The top spectrum was taken by a 300-mm Czerny–Turner spectrometer with spherical mirrors. The grating is also 1800 mm^{-1} with a RD of 1.34 nm/mm ; the resolution is 60 pm with slits $10 \mu\text{m}$ wide and 4 mm high (10,000). The light source is a standard Pen-Ray® Hg calibration lamp. The central peak represents the resolution of the monochromator. Almost all of the side lines are covered by the envelope; only the small peak on the left, -22.5 pm from the center, appears as a shoulder.

2.6.6 Image quality: Q-factor or fidelity

Every optical system suffers from the competition between light flux and image quality. The light flux increases with the illuminated size of the optical components, but the image quality degenerates at the same time. There is no general equation to calculate the image quality, but ray tracing software can put the complex pathways of rays together and display the expected image distribution. There is a general rule stating that the larger the focal length is, the less the impact of irregularities on the surface of optical components. The correlation is linear; larger diameters produce higher probabilities of beam deviation, which leads to aberrations. This correlation is typically a square function. The more the reflection angles differ from the “normal,” the stronger the corrections required in the optics, leading to a higher probability of aberrations. In both cases, the chance of rays reaching the target without crossing the axis of symmetry (horizontal and/or vertical) on their way from the entrance to the exit increases.

Defocusing effects are the result. At the slits, the quality of the edges and the symmetric movement of the jaws play a key role. Imperfect edges will diffract and scatter the beam, and asymmetric movement will force the rays out of the center. All of these parameters affect the image transfer, which in turn influences the correct reproduction of the spectral signal shape. All kinds of defocusing will reduce the signal concentration in the output. Besides the loss in resolution, the contrast suffers, as well. In principle, a well-designed spectrometer can be tuned for certain parameters—such as the best image over an area, best resolution, best contrast, and so on—but not for all at once. The general estimation of aberrations, creating an additive distortion to the focus, is possible by a group of equations. The real aberrations depend on a number of parameters and the quality of the components used. It is due to physics that aberrations are generally small in the center of a field. Note that unless marked otherwise, estimated values are valid in the outskirts of the field area; all calculations are valid only for estimation.

2.6.6.1 Calculating aberrations

The basic aberration equation reads as follows:

$$O_1 = W \times m_s/f^2, \quad (2.26)$$

where O_1 is the aberration, the added distortion of optical information in a single-axis instrument (the system shown in Fig. 2.1 is single axial because the deflection originates from dispersion); W is the active width of the grating or prism, or the illuminated width of the optical components; m_s is the actual slit width of a monochromator (in a spectrograph, the used width of the detector field is to be applied); and f is the focal length when focused beams are involved (for collimated beam areas, the distance between the two surfaces is the valid value).

2.6.6.2 A reflecting spectrometer has two or even three axes

As mentioned before, increasing the internal angles (ϵ , δ , and ν) leads to increasing aberrations. Equation (2.27) provides a multiplication factor that results from the internal angles:

$$E = O_1 \times (1 + \sin \delta), \quad (2.27)$$

where E is the deformation in a dual-axis spectrometer, and δ is the internal full-opening angle at the component with the widest angle, such as the ν angles at a mirror or the δ angle with respect to ϵ at a grating. The grating's working angle ϕ plays no role because it is required only for dispersion calculation. However, coma (Section 2.6.7.7) originates from the working angle of a grating or a prism.

This calculation is performed with the component that is illuminated by the widest angle in the chain. Note that the estimation is only valid for spherical optics. Instruments with image correction, such as toroidal mirrors, need different and more complicated algorithms.

2.6.6.3 The slit height also influences the total aberrations

The following equation is used to estimate aberration with curved slits:

$$H = E \times (1 + h/f), \quad (2.28)$$

where H is the total aberrations, including the vertical factor in systems with curved slits (it increases the distortion of the horizontal aberrations, and it originates from the vertical dimension of a curved slit or detector), and h is the illuminated slit height (curved slits only).

The estimation of spectrometers with straight slits is based on the radius that a curved slit would have, if it were used. It is the lateral distance of the slit center to the center of the instrument. The resulting additive aberration is only valid for systems with straight slits. This value instantaneously appears in the middle of the horizontal spread as soon as the vertical position departs from center.

The following equation is used to estimate aberration with straight slits:

$$O_{ss} = E + (h^2/r), \quad (2.29)$$

where O_{ss} is the minimum aberration in a dual-axis instrument (such as a reflecting spectrometer with straight slits), r is the radius of curvature of the slits (identical to the distance slit to the center of the instrument and valid only

for symmetric systems without imaging correction. For instruments with aspherical optics, a ray-tracing program is required), and h is the slit height.

Spectrometers with imaging correction might have stronger distortions in the center of the field but much better values in the outskirts of the xy frame. Figure 2.22 in Section 2.6.6.5.2 illustrates some situations graphically.

2.6.6.4 The quality factor

As with other specifications, the system resolution is the result of several parameters. The general expression for functions such as that is

$$\sigma_{System}^2 = \sum_{\text{all sources}} \times \sigma_i^2, \quad (2.30)$$

which means that the square root of the sum of all squared tolerances will result in the technically possible final parameter, here the resolution. Division of the technically realized value by the theoretical optimum of a parameter leads to a result between 0 and 1, which is called the parameter's "Q-factor" or "fidelity."

In case of the image transfer, the factor is called Q_i , the quality of imaging. Calculation takes the aberration in the output relative to an entrance spot with a 100-μm diameter. If, for instance, the aberration is up to 300 μm, $Q_i = 100 \mu\text{m}/300 \mu\text{m} = 0.3$. Table 2.1 lists typical values for Q_i found for spectrographs with internal imaging correction.

The Q_i value might be of no experimental importance to monochromator applications; however, considering that internal imaging correction compromises the x resolution (wavelength), it might be an interesting point. To provide a choice, some manufacturers deliver spectrometers with spherical mirrors instead of toroidals upon request.

2.6.6.5 Image-transfer issues

Light rays in optical systems concentrate in the center of the traveling channel, after they have passed through fibers, mirrors, or lenses. The more components present in the chain, the stronger the distribution of light from the center to the radius of the optical components.

A centrally focused spot on the entrance slit of a spectrometer will, in most cases, have its highest intensity in the center. It follows that the collimator and the

Table 2.1 Quality-of-image factors for various parameters.

Focal Length	Aperture	Estimated Q_i
150 mm	f/4	0.250
250 mm	f/4	0.275
300 mm	f/5	0.300
500 mm	f/7	0.500
750 mm	f/9	0.650
1000 mm	f/10	0.830
1500 mm	f/15	0.950

grating also have the highest light density in the center, as well as continuous decay to the radius. A rule of thumb says that the center 30% of a grating area carries ~50% of the signal, and the center 50% of the area carries 75–80%. Taking into account disturbances such as aberrations, astigmatism, and coma, the optical aperture grows wider as the light intensity increases. This means that in surface areas where intensity slowly drops, disturbance quickly increases. The wider the allowed angles in the system, the more rays will travel without crossing the center but still reach the target. These “off-axis rays” are the largest cause of troubles.

2.6.6.5.1 *Image aberrations*

The off-axis illumination of optics opened in the plane of spectral dispersion, called the “horizontal” plane, results in two effects (see Fig. 2.21):

1. The focus of the horizontally oriented rays is closer to the focusing mirror than the focus for the vertically oriented rays. The beam has an elliptical shape, turning between the two planes. For pure spectroscopy, a vertical aberration is not a problem so long as the horizontal-oriented information is sharp; that plane represents the tangential focus T . The best vertical focus is important for multi-stripe spectroscopy and imaging. This plane is the sagittal focus S . For imaging spectroscopy applications, a compromise between both planes can be chosen: M . The planes are often tilted against each other, depending on the optical layout of the instrument. Applying nonspherical optics can strongly reduce the distortions.
2. Guiding light to mirrors or lenses out of axis will result in different travel pathlengths of rays, which leads to a widened focus. The rays do not pass through the same spot; this effect, called “coma,” results from moving the angle of a grating or prism when changing the wavelength. The coma effect can also be reduced by nonspherical optics, which is often optimized to a certain angle of the dispersing element.

Both effects increase with aperture, typically with the square of the change. A straight transmission system with lenses is easy to understand. The more a ray travels away from the center axis, the longer its pathway to the target. Finally, the rays running close along the center axis create a sharp image in both the x and y axis. The rays that travel further and experience stronger

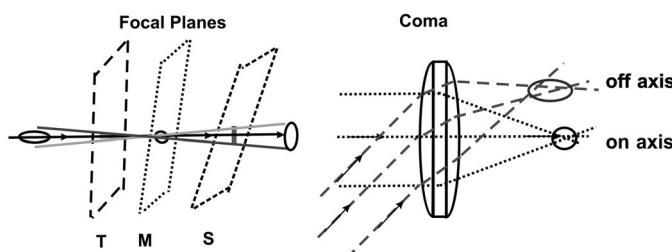


Figure 2.21 Common distortions (aberrations) in spectrometers with laterally opened angles.

changes in direction will have a greater variety of possible landing positions. In other words, they will have a wider spread at the target, which means nothing else than defocusing. A narrower optical channel produces a sharper focus but also reduced light flux. In a reflecting system based on mirrors (which is true for almost all spectrometers), angles can be found in at least one plane. Consequently, even in the center of the target field, rays travel different pathways: the vertically oriented rays differ from those horizontally oriented.

A classical spectrometer therefore has two focal planes: one for the vertical (sagittal) orientation, and the other for the horizontal (tangential). In a monochromator, the main interest is the tangential focus, representing the wavelength axis, whereas the sagittal focus is of similar importance in imaging spectrographs. If, in a monochromator, the sagittal focus is bad but all photons reach the single-point detector, then all is well. Therefore, monochromators (with tangential optimization) use spherical mirrors for the best wavelength resolution and accept a defocused height. Spectrographs, on the other hand, incorporate correcting optics to reach a compromise. (The methods of correction are further treated in Section 4.1.6.)

Historically, Littrow and Ebert–Fastie spectrographs with photographic plates or films were popular for emission spectroscopy, comprising the entrance and exit below and atop of (respectively) the dispersing element. These instruments were equipped with spherical optics, providing reduced sagittal aberrations at the expense of spectral resolution. Because the best focus in the output of a spherical spectrograph is at a radius, signal recording with curved films has been very popular. Films up to 25 cm in length recovered large spectral intervals with excellent resolution.

Figure 2.22 illustrates two kinds of ray transfer. The upper box represents the entrance illumination, and the lower box is the exit function, related to each of the four kinds of entrance illumination, thus showing the different transfer behavior. The figure demonstrates how the image of the related input signal will probably appear. Leftmost of the three horizontal versions is the optimum stigmatic transfer function by an Ebert–Fastie or symmetric Czerny–Turner, working with curved slits or with fibers on a curved radius. In the center is a toroidal-corrected Czerny–Turner for imaging applications. Rightmost is an uncorrected Ebert–Fastie or Czerny–Turner with straight slits and spherical mirrors. Each of the three blocks contains four modes of illumination. Leftmost involves a transfer slit to a slit at full height. Second from left is a number of spots like those at the end of discrete fibers over full height. Third from left is the transfer with a rather short slit height. The rightmost shows a single spot, such as a laser spot or a single fiber, in the center of the entrance.

2.6.6.5.2 Astigmatism

Astigmatism adds to the aberrations already discussed in the context of spherical mirror systems. If the height used in the entrance increases, the rays travel longer distances from the extremes in the entrance or exit positions to

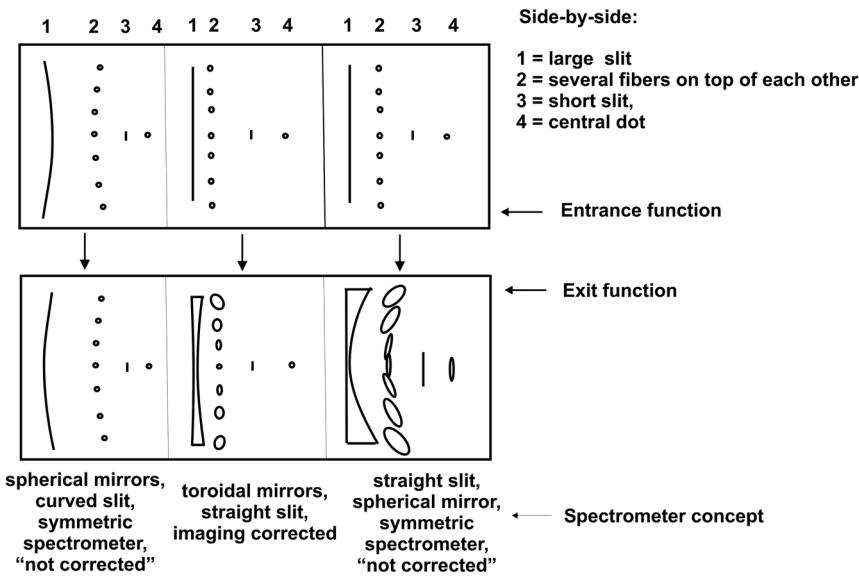


Figure 2.22 Stigmatic and astigmatic ray transfer for three typical systems.

the grating, compared to the rays traveling from center to center. This is true for both planes. If only one wavelength, such as a laser or an atomic line, enters, then the monitored signal in the exit will appear with barrel- or cushion-shaped deformation. In a symmetric spectrometer, the pathlengths can be made uniform by curving the slits with the radius of their distance to the center line of the spectrometer. This situation is illustrated by the leftmost pair in Fig. 2.22. All rays of the same wavelength in the tangential plane have the same travel length, and they arrive at the exit with good spectral and spatial resolution.

This phenomenon is the reason why spectrometers with straight slits are most often not specified with the full slit height. The customer should expect the height applied for the specification to be mentioned. Symmetric instruments with curved slits can be used at full height without sacrificing resolution. In an exit field, the rays of one wavelength will appear well focused but on a crescent-shaped line instead of a vertical line. It will be difficult to sum up pixels of the same wavelength. Also difficult is the wavelength-dependent assignment of arrangements with multiple spectra in vertical order.

2.6.6.6 Spectrometers with internal image correction

Toroidal mirrors or a Schmidt plate compensate for both aberrations and astigmatism, but at the expense of the pure monochromatic function (resolution). They offer a compromise that is not optimal for either the tangential or the sagittal plane, but it is an acceptable solution over a defined image area (see Section 4.1.6.2). The wider the aperture is in relation to the focal length, the worse the correction in the outskirts of the field area.

The larger the lateral distance is between the entrance and exit, the stronger the aberrations and the more difficult the correction.

2.6.6.7 General aberrations and coma

All deviations from the ideal transfer are categorized as “aberrations.” Negative impacts are produced by asymmetric moving slits, low-quality slit jaws, the accuracy and quality of the mirror and grating surface and structures, wide angles between the components, and a wide aperture. All kinds of angular increases lead to aberrations, one of which is the coma effect. All aberrations can be pre-defined by ray-tracing software during the design phase. In most cases, ray tracing matches reality very closely.

Coma is a defocusing effect that originates from the beam transfer between mirrors (primarily the collimator and grating). The larger the optical components are, the more rays diverge and do not travel along or around the center axis of the optical channel. They appear “smeared” in the exit. It is possible to shape the mirrors such that, at a certain grating angle and with spot illumination of the instrument, the median path of all tangential rays is the same. This arrangement will result in a sharp, well-focused wavelength signal in the exit. If no special grating angle is designed, this scenario is achieved if $\alpha = \varepsilon_1$ or if the collimator and grating nearly face each other. As the grating turns, the peripheral rays become unfocused and land single-sided beside the target in the exit. The image of a spectral point now looks like a comet with its tail, which is where the name “coma” comes from. The coma effect increases as the grating angle increases. When moving the grating axis N to the opposite side, the coma will also change direction. It is not possible to pre-define the direction of the coma without knowing all angles inside the spectrometer, but it is a reproducible and traceable effect. Unfortunately, there is no other way to reduce the coma other than limiting the aperture. As shown in Fig. 2.23, the aberration originates mainly from the extreme surface area of mirror and grating; thus, it might help to put an asymmetric mask on the mirror, shading only one side. This technique normally works better than circular masking and consumes far fewer photons. If only a short angular grating change is required to perform the measurement, a small mask might be sufficient. Shading one-fifth of the mirror surface will probably produce satisfying results, consuming perhaps 10% of the light flux.

Figure 2.24 provides an example of spectral falsification produced by coma, demonstrated by two nearby peaks, both of which are coma shifted towards the blue. Coma affects the resolution and amplitude. The correct curves are dotted, whereas the solid curves are coma-shifted to the left. The dotted green curve is the correct result, compared with the faulty solid-red curve.

Once again, there is competition between focal length and aperture. If the experiment in question does not ultimately require the last photon, it might be better not to chase it. The upshot is that it is always possible to reduce a large aperture by masking.

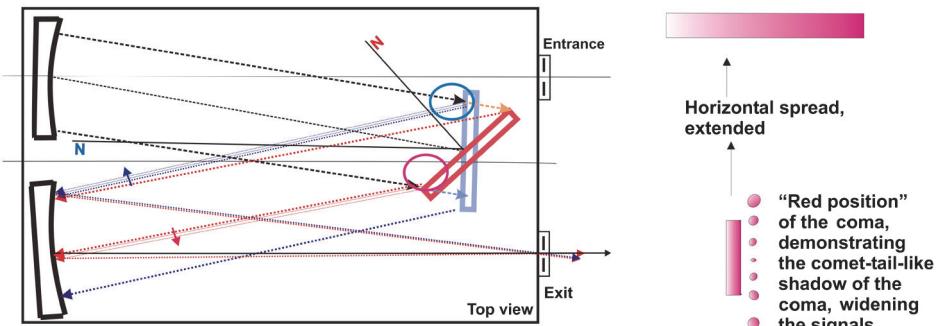


Figure 2.23 The coma originates. The blue mark applies if N is “inside” the collimator; the red mark applies if N is “outside.” The arrows show the direction of the coma’s spread.

2.6.7 False and stray light, and contrast

The unwanted effects of false light and stray light have different origins, and they are therefore treated separately.

“False light” is light in the exit of a spectrometer, which can be inhibited. This term includes

- unwanted spectral orders resulting from the diffraction grating (they can be filtered out),
- “grating ghosts” coming from nonhomogeneous structures of the grating (they can be inhibited by the use of better gratings or masking the bad areas of the grating), and

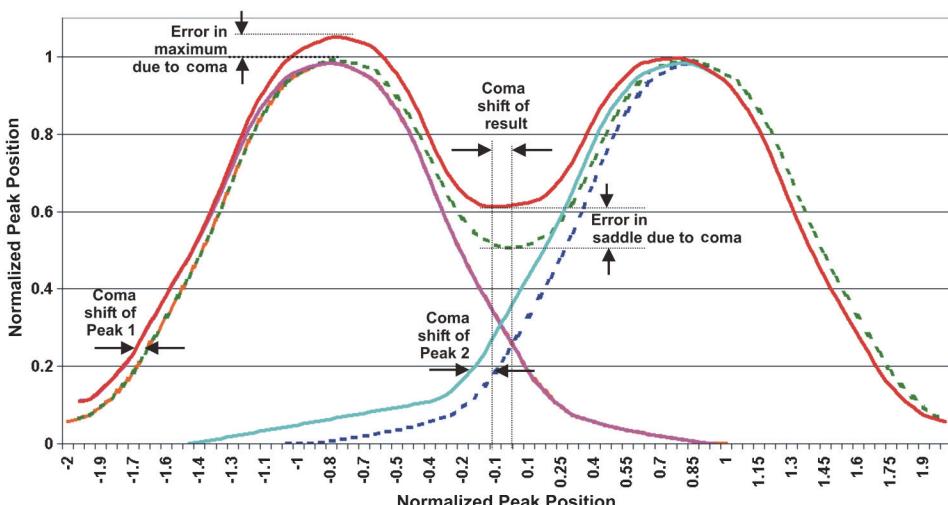


Figure 2.24 Example of spectral falsification from the coma effect.

- environmental light coming from the experiment or the surroundings, such as lab lamps. It often enters the spectrometer in strange ways and angles, but contributes to the measurement. It can be blocked.

Minimizing false-light problems is rather easy once they are detected.

“Stray light” originates in the spectrometer itself. It cannot be completely inhibited, only minimized. This term includes

- internal reflections, such as rays that leave the proposed optical channel and reflect from surfaces such as walls (even the “blackest black” still reflects about 6%);
- rays leaving the proposed beam and hitting edges and substrate areas of the components (e.g., an overilluminated mirror or grating);
- multiple-dispersed light coming from multiple routes between the mirror and grating. Such rays can stay in the beam but arrive at the exit at wrong places (wavelength or geometrical position) and overlay the measured signal (not to be confused with unwanted orders);
- light scattered at imperfect surfaces of the mirror or grating, reaching the output;
- unwanted diffraction of light (sources include slits of low quality and other sharp edges. Other phenomena result from extreme, yet allowed, angles inside the beam skeleton, such as coma); and
- unexpected wavelength ranges. The inner walls, component holders, etc., can be tuned for the best stray-light behavior in certain wavelength ranges. For instance, a coating can be optimized for the visible range, but it can be a rather good reflector in the IR or fluorescent in the deep UV. It is hard to optimize for low stray light from the VUV to the far IR.

2.6.7.1 Reducing stray light

Improved surface quality of components reduces scattering. The frames, flanges, and edges of components should not be polished, blank, or sharp-edged to avoid reflection of arriving (stray) light. The inner volume of a spectrometer should be large because stray light is reduced in an exponential manner with volume. The application of masks and baffles can help channel the beams and reduce scattered light. If possible, only the required wavelength interval should enter the spectrometer.

UV light creates scattered signals that can arrive at the detector even if no UV is required for the analysis. Because scattering grows with eV^4 , a signal at 250 nm creates ~250 times the scatter compared to a disturbance originating from 1 μm . Infrared light, in turn, is heat, which influences the thermal stability. The further the entering spectral interval reaches into the IR, the more heat it introduces. All kinds of distortions apply if they reach the output. The interior of a spectrometer should be as empty as possible. The surfaces of walls and covers should be rough to reduce direct reflection. If constantly active electronic elements (which create heat or light) are located in the optical

room, the apparatus might not be the best available choice. Modern systems incorporate stepper motors that can be mounted externally. The supporting electronics should sit in an isolated part of the housing.

Chapter 8 addresses the issue of stray light in further detail.

2.6.8 Contrast ratio C

The contrast C , which describes the useful signal versus the destructive signal, strongly depends on the application. If only discrete wavelengths or a narrow interval (as from an LED or diode laser) enters, a good contrast ratio can be expected. There will be almost no vagabond energies. If, on the other hand, the spectrometer is “filled” with broadband light, the contrast will surely suffer. Towards the UV, scattering effects increase with an exponential power of 4. Stray light spreads in all directions and will “get lost” much quicker in a larger instrument. Even if instruments with identical apertures but different focal lengths are compared, the longer system will have a distinct advantage in contrast!

A monochromator, compared to an identical spectrograph with a 25-mm-wide output field, has an advantage in contrast by a factor of 10–100. An important stop for stray light is the monochromator’s exit slit. It presents a solid obstacle for rays under angles outside the defined range. A spectrograph, in turn, has a large field at the exit that is open for a wide angular range of exiting rays. Vagabond rays have far better chances to reach the detector that way.

Thus, once again focal length and aperture compete against each other. Although the light flux L theoretically drops with the square function of the aperture n , the contrast C rises faster than squared, superseding the light loss. Section 2.6.2.3 shows that the experimental light flux L is far better than theoretical expectations would suggest. Again, focal length dominates aperture, supported by the fact that it is much easier to apply baffles and light stops in larger spectrometers than in smaller ones.

2.7 Mechanical Stability and Thermal Influence

Mechanical stability and thermal influence are often underestimated parameters. Instability can result from construction, based on materials not equally responding to thermal changes, production quality, and mechanical stability. A likely source of troubles is vibrations penetrating from the environment or from the system itself; for example, a stepper motor creates strong torque moments. While moving the grating or a different device, vibration or resonance might need several milliseconds to diminish after actuation. A stepper motor system with solid coupling will introduce “a damped modulation” to the grating. Data collection should wait until the grating ceases to vibrate, otherwise the signal can become frequency (wavelength) modulated. A system with a soft coupling, on the other hand, will need some time until the grating has reached the addressed position. Again, a short period before taking data is recommended. Hysteresis (backlash) in an actuation system generally cannot be avoided

completely. The grating positioning system should always approach from the same side. There are several steps a designer and programmer can take to keep instruments stable and useful. Unfortunately, the customer can almost do nothing to address those troubles if they are evident. In so-called “price worthy” spectrometers, these impacts are found more often than one might think.

Material-dependent torsion will become detrimental as the system experiences thermal changes. Thermal drift, by nature, cannot be avoided completely and is most recognizable in large systems, but it should have linear and reproducible effects that leave neither twisting nor permanent residuals. So long as a strong (thick), homogenous baseplate carries the components, torsion and twisting should not occur. If all of the components sit on the baseplate, the housing should only be a hood with no real carrying function, in which case the hood can be of different materials. Precautions against mechanical and thermal stress should still be taken: if the housing also carries optomechanical parts, it should be of the same material as the baseplate to ensure homogenous drift behavior and provide high mechanical stability at the same time. A very elegant but expensive way involves mounting all important components to a cage of Super Invar® rods and covering it with a housing that can be made of sheet metal or even a synthetic material, so long as it is light, tight, and does not stress the rod cage thermally or mechanically.

2.7.1 Measuring thermal variations

The following equation can be used to calculate thermal variation:

$$dx = K \times 2f \times dT, \quad (2.31)$$

where dx is the change in focal position after a thermal event (it is the distance from the mirror to the entrance and the mirror to the exit. The distance to the grating is negligible because of the collimated beam), K is the thermal coefficient of the materials in kelvin^{-1} , f is the focal length in meters, and dT is the thermal change in kelvin.

2.7.2 Defocusing effects

The impact on the focus, and thus the resolution, is combined with the light concentration in the output. It is described by

$$dy = dx/n, \quad (2.32)$$

where dy is the increase of the focal area after the thermal change, and n is the f -number n that is actually used.

2.7.3 Typical thermal constants

Most spectrometers are made of aluminum and completed with steel, glass, brass, or bronze. In astronomy, Zerodur® is applied, and some rare, large spectrometers incorporate Super Invar®. Table 2.2 shows some typical thermal coefficients K between 15–30 °C.

Table 2.2 Common materials and their thermal coefficients.

Material	K at Ambient
Aluminum	22×10^{-6}
Steel	11×10^{-6}
Glass	9×10^{-6}
Zerodur®	0.03×10^{-6}
Super Invar®	0.2×10^{-6}

Consider two examples involving standard spectrometers: a 0.3-m monochromator with $f/5$, and a 1.0-m instrument with $f/10$. Both are made of aluminum and have 20-mm-thick glass optical components. The thermal step is 5 K.

Example: $f = 0.3$ m

$$dx_{0.3m} = 22 \times 10^{-6} \times 600 \text{ mm} \times 5 \text{ K}/K = 66 \mu\text{m}.$$

From this value, the extension of the two mirrors needs to be subtracted (they move in the opposite direction):

$$dx_{\text{mirror}} = 9 \times 10^{-6} \times 40 \text{ mm} \times 5 \text{ K}/K = 1.8 \mu\text{m},$$

leading to a total $dx = 64.2 \mu\text{m}$ and $dy_{0.3m} = 64.2 \mu\text{m}/5 = 12.84 \mu\text{m}$.

Thus, the 0.3-m unit will defocus by 13 μm . If the slit width was set to 20 μm and the expected resolution was reached in front of the thermal step, the performance reduces to 2/3. As a consequence, the intensity in the output will drop to 40% of the original value because $\sim 1/3$ is lost at each slit.

Example: $f = 1.0$ m

$$dx_{1m} = 22 \times 10^{-6} \times 2000 \text{ mm} \times 5 \text{ K}/K = 220 \mu\text{m}$$

is reduced by 1.8 μm , resulting in $dx = 218.2 \mu\text{m}$ and $dy_{1m} = 218.2 \mu\text{m}/10 = 21.8 \mu\text{m}$. The long monochromator will experience a change in focal length of 218 μm and will defocus by 22 μm . With a slit width of 20 μm , the resolution will degrade to less than half, and the light flux in the output will drop to 20%.

In addition, the change in dimension will affect the grating angle, thus influencing the calibration, dispersion, and (in spectrographs) image quality. Because those parameters depend on the specifics of the setup and driving system, no calculation base can be provided.

Thus, it is worthwhile to monitor the thermal environment. If no stability can be realized, it is a good idea to calculate the stability before measurements encounter errors. Thermal control of the spectrometer will eventually pay off. The interior of a spectrometer should not contain components that distribute heat because they can interfere with the thermal balance of the system.

2.7.4 Minimizing environmental influence

If vibrations are created in the surroundings, the spectrometer can be set on a vibration-damping table. Most spectrometers come with rubber feet, but even though they provide a damping effect, they might be undesirable when sitting on an isolated table. It might be better to remove the feet; however, it might be counterproductive if the vibration created by the spectrometer's stepper motors is transferred to other components on the table. Other options involve using feet with internal damping or mounting the spectrometer in a housing or skeleton with holding springs. A closed housing can be useful when combined with controlled temperature, and it can be purged. Before buying, check if the system is constructed for stability and ruggedness, and can carry the additional load from extra components (lamps, filter wheels, camera, cooled detector housing, etc.) without drifting over time.

2.8 Reduction of Unwanted Spectral Orders, and Other Filtering

The wavelength at the exit of a grating spectrometer results from the product of $m \times \lambda$ in the grating equation. If the result can lead to spectral overlap and obscured data, precautions must be followed to avoid them.

2.8.1 Long-pass filters

If only the first order is used, which is most often the case, edge filters (long-pass filters) are sufficient. They block shorter wavelengths that appear and overlay in higher orders. The supplier's information about edge filters defines the nominal wavelength in most cases as the center wavelength of the rising edge, i.e., the 50% mark. This number is valid for collimated illumination. The difference between rise and the maximum transmission depends on the material and production process. If no curve is available, a wavelength difference of about 5% of the center wavelength can be applied. For a nominal wavelength, e.g., 570 nm, the rise can start at 555 nm (which is the 5% T-value) and reach saturation at 585 nm. With increasing illumination angles, the edge curve becomes more moderate. An $f/4$ cone (~ 15 deg) produces a curve that will be $\sim 7\%$ wide. Filters for the UV range are shallower and sometimes show some modulation. In practice, most filters in spectroscopy are not illuminated in parallel, which leads to a red-shifted maximum. Figure 2.25 shows typical curves in the UV–Vis range.

Thus, the 570-nm example filter might be suitable for use above 600 nm. For the same reason, and to provide some flexibility, the steps between filters should leave some overlap. A typical filter chain for a system between <200 and >2800 nm would be: none, 320, 530, 820, 1100, 1800 nm. An additional beam stop is a good option to allow background measurements at any time. Long-pass filters for wavelengths >1000 nm have a limited working range. They close down again at high wavelengths and can project side transmission at remote spectral distances.

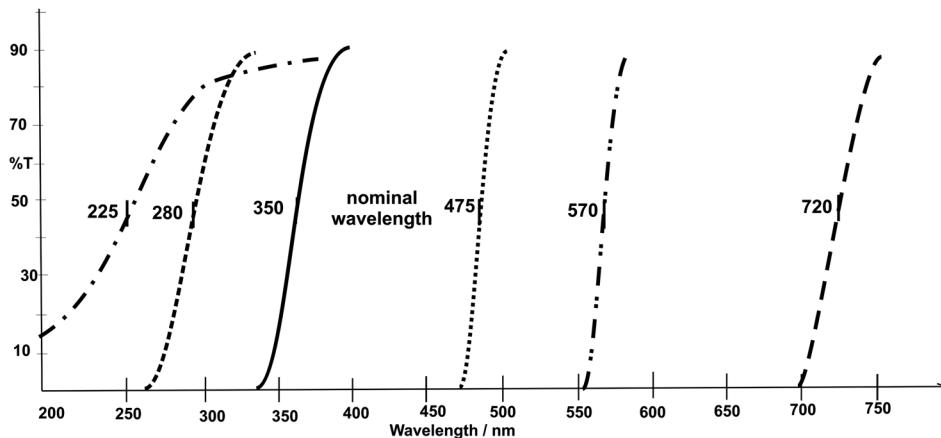


Figure 2.25 Typical order-sorting filter curves (long-pass filters) in the UV–Vis.

2.8.2 Band-pass filters and prism

A spectrometer used in different diffraction orders will optimize resolution, dispersion, and efficiency if the grating is prepared for it; however, this complicates the application. The orders above and below the actual one might need to be reduced if they contain injurious light. Two methods can cover the task: band filters or a prism pre-disperser. A variety of band filters are available for all wavelength ranges. The bandwidths will reach from 10 nm FWHM to about one-fourth of the central filter wavelength, as shown in Fig. 2.26. The wider the filter's transmission range is, the higher and flatter the curve, which can vary between 10–90% T . Optimized band filters are also available for certain applications. Finding a suitable band filter requires evaluating the position and bandwidth, as well as the rising and falling edge.¹

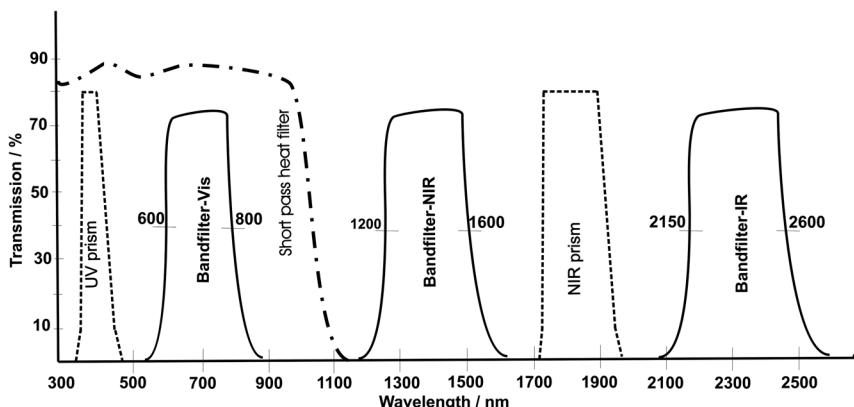


Figure 2.26 Typical transfer curves of band-pass filters, short-pass filters, and prisms.

It might not always be possible to cover one full spectral octave (one free spectral range, or FSR) with a single filter because the filters might be too narrow. Further complicating matters, due to Eqs. (2.2) and (2.3), the FSR will become smaller with increasing order numbers. If wide spectral ranges need to be covered, filtering can become a high-cost factor in the system.

An excellent alternative is the prism pre-monochromator. It can be built in such a way as to not limit the aperture of the main spectrometer and still provide a spectral band at the main entrance wide enough to supply the requirement but narrow enough to remove spectral-order overlap. In Fig. 2.26, two prism examples are represented by curves at 400 and 1800 nm. The function of a prism is based on the change of refractive index and the changing working angle versus wavelength. The transmitted bandwidth in the UV is much smaller than in the red. For most applications, that behavior is advantageous, and prism filters provide further benefits: the transmission is very high, they increase the total dispersion, and they reduce the stray light.

It is always a good idea to not transfer any more light into the spectrometer than what is required for the application because excess light can create distortions, reflections, and heat. Therefore, filters and other pre-selectors should be located in front of the spectrometer instead of behind. Another alternative for pre-dispersion is a grating pre-monochromator running in first order. If the order calculation fits and its efficiency curve provides enough signal over the target range, it might be better and cheaper than a set of filters.

A prism or grating pre-selector should be designed like any other spectrometer. Its output slit will be the entrance slit of the main stage. Keeping the rule of optical transfer [Eq. (1.11)] in mind, it can be used to fit the aperture of the illumination beam and spectrometer.

2.8.3 Short-pass filters

Using broadband light sources often provides a light interval much wider than required. Strong light >1300 nm can create thermal changes in the instrument, the consequences of which are discussed in Section 2.6.9. It can be beneficial to block long wavelengths with a short-pass filter, as the third curve in Fig. 2.26 shows; unfortunately, they also have a transmission limit towards low wavelengths, acting more like very wide band passes. Water is a potential heat filter because above 1150 nm, it has many strong absorption bands. A transmitting cylinder with distilled water between the lamp and spectrometer makes for an efficient heat stop. Note that heating water will create turbulence, and interference can follow. Circulating the water will prevent the turbulence, but even the laminar flow can interfere. Another possibility is the series combination of long-pass and short-pass filters. However, care must be taken to avoid interference and reflections between the surfaces. This solution also needs parallel light.

2.8.4 General filtering techniques

Interference filters work in accordance to the Fresnel laws, which describe transmission and reflection at optical interfaces. Interference effects are used to create filters that transmit one or several defined bands (band-pass filters), or cancel defined bands (band-trap or notch filters) out of the transmitted spectrum. The interference can occur internally at the interfaces or use additional thin-film technology at the surface. Interference filters require parallel light for proper functioning. Absorption filters have either absorbing material added to the substrate, or they are coated with one or more layers of absorbing material. They are available in long-pass and short-pass versions. Within certain limits, they work with parallel, converging, or dispersed light. Color filters work with complementary colored substrates. Whereas the other filter types might not reveal their function to the human eye, color filters allow estimation of their working range. Color filters are not very sensitive to the illumination angle. Further details are described elsewhere.¹

2.8.5 Notch filters

An inverted band filter functions as a notch filter. One of the band filters shown in Fig. 2.26 is a band notch, which is used in reflection instead of transmission. By nature, this arrangement is not ideal, but it is also seldom required in spectroscopy. It is more often the case that a very narrow notch is wanted, in order to block a single spectral line and let the rest pass. By 3D holographic application, extremely narrow banded notch filters can be produced. They only cut a very narrow band out of the spectrum. That allows one to stop a single spectral line, which is helpful in applications such as fluorescence and Raman spectroscopy. The technique is described in Section 9.3.

2.9 General Collection of Performance Parameters of Spectrometers

This chapter has presented all of the parameters required to enable the general characterization of spectrometers. Figure 2.27 combines the typical behaviors. The vertical axis is symbolic of each of the parameters, but linear. The theoretical limit is “max,” which in reality cannot be reached by any of the parameters.

The parameters shown in the figure are as follows:

- The **R_r/R_p** dotted curve describes Q_r , the ratio of the real resolution to the theoretical value of the grating’s resolving power. If, as normally happens with an extended focal length, the width of the grating increases, R_r and R_p will also increase, thus improving the real resolution even if Q_r remains similar.
- The **C** dotted curve represents the contrast, i.e., the signal-to-background ratio.

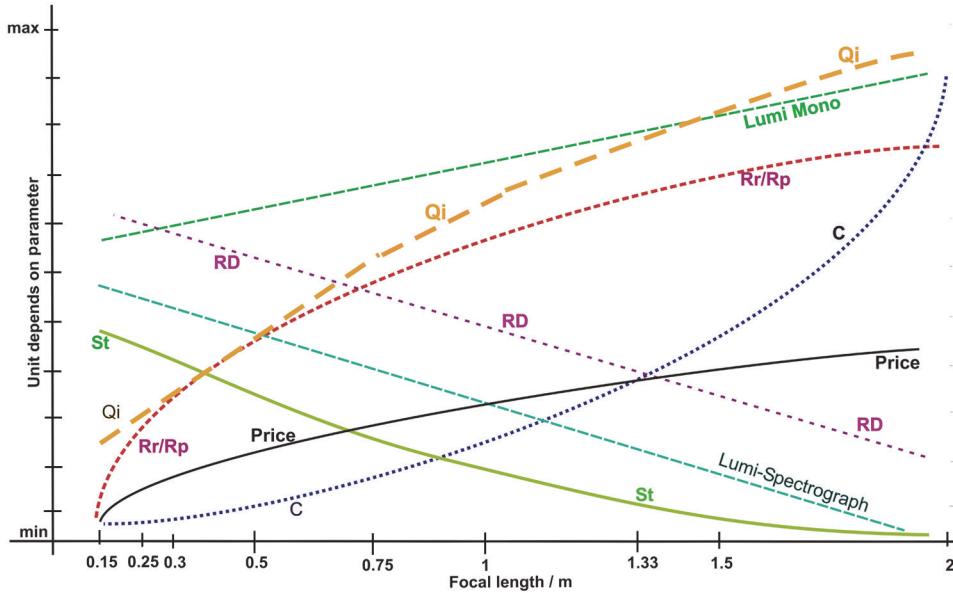


Figure 2.27 Collection of the most important spectrometer parameters.

- The **RD** dashed line indicates reciprocal dispersion.
- The luminosity (light flux) is separated in two curves.
 - The **Lumi Mono** dashed curve adjusts the slit width and height for the required beam and resolution, leading to improved values with increasing focal lengths.
 - The **Lumi-Spectrograph** dashed curve implements a slit width and height fixed to the size of the detector pixels or fixed bandwidth. This naturally leads to a decreasing L with longer focal lengths.
- The **Qi** dashed line is the quality factor of a spectrometer for the image reproduction. It describes the surface quality of the components in relation to the focal length and aperture, as well as the reproduction quality, depending on the internal angles.
- The **St** curve demonstrates the general thermal stability. If the system is thermally stabilized, this value becomes constant.
- The **Price** solid curve is the general behavior of cost. It shows that the cost increases distinctively slower than spectral performance increases.

Reference

1. Oriel Instruments, Inc., *The Book of Photon Tools*, Stratford, CT (2000).

Chapter 3

The Dispersion Elements: Diffraction Grating and Refraction Prism

3.1 Introduction

The equations required to specify a spectrometer [Eqs. (2.1)–(2.5)] are discussed at the beginning of Chapter 2. This chapter mainly deals with the specific features of the different types of gratings. Information brochures and books are available from some grating suppliers.^{1,2} The parameters of a prism are also discussed.

3.2 Diffraction Efficiencies and Polarization of Standard Gratings

If a holographic grating were kept unblazed, the orders $+d_1$ and $-d_1$ combined will contain $1/e$ ($\sim 68\%$) of the dispersed light of the first-order wavelengths. Therefore, the “right direction” $+d_1$ receives, at best, 34% of the available light of the wavelength set. $+d_2$ and $-d_2$ combined contain 68% of the rest (which can theoretically be up to $\sim 22\%$) of the available light of that wavelength, and so on.

Blazed gratings, however, combine the effects of diffraction and (micro) reflection. As Fig. 2.2 shows, the lines are sawtooth-shaped. At the blaze angle, the incident light fulfills the reflection law; now, up to 85% of the available photons at a given wavelength can be dispersed in the “right direction.” The diamonds of ruling machines are shaped and adjusted for the best structure in order to achieve this behavior.

Blazed gratings show the maximum diffraction efficiency at a certain angle α in relation to the incidence beam I. The manufacturer documentation of a grating typically specifies the blaze angle and a wavelength resulting from that angle. If the specification is not marked to be valid for a certain spectrometer setup, it refers to the first order and the Littrow condition.

Littrow (see Section 2.3.1) means that I and d , and therefore α and β , are identical. In most current spectrometers, that is not the case. Consequently, the grating data of actual spectrometers can slightly differ from the data sheet. Because a blazed grating is optimized to a certain angle, light from the first, second, third, or even higher orders will be diffracted at similar efficiency under the blaze condition. In a measurement, this rule does not always hold up; with the higher orders, the dispersion is also multiplied [see Eqs. (2.4) and (2.5)]. In most applications, this will produce a signal at the detector whose magnitude is divided by the degree of the order. Figure 3.1 illustrates the typical behavior of a grating as a function of the plane of polarization. The solid center curve is the main curve, showing the average 45-deg polarization. The bottom dotted line represents the typical efficiency of the parallel P plane. In relation to the grating, it is the plane that is oriented parallel to the lines. Physically, the plane will rotate if the spectrometer (and therefore the grating) is tilted. The top dashed curve represents the S plane, which is perpendicular to the grating lines.

Below the blaze angle, the P plane is generally stronger, whereas at higher angles the S plane is stronger. This effect can be used to either select the grating during the experiment or arrange the light to enter with the better polarization. If the light enters a spectrometer perfectly scrambled, it will almost always leave the system partly polarized. Finally, it is important to point out that the behavior shown in Fig. 3.1 is only for general information. Gratings have very different efficiency and polarization characteristics! Even if several types with a common line frequency are compared, significantly large differences might be evident. Planning a spectrometer means collecting the actual data of the gratings under consideration and interpreting them. In general, for standard ruled or holographic gratings, it can be assumed that

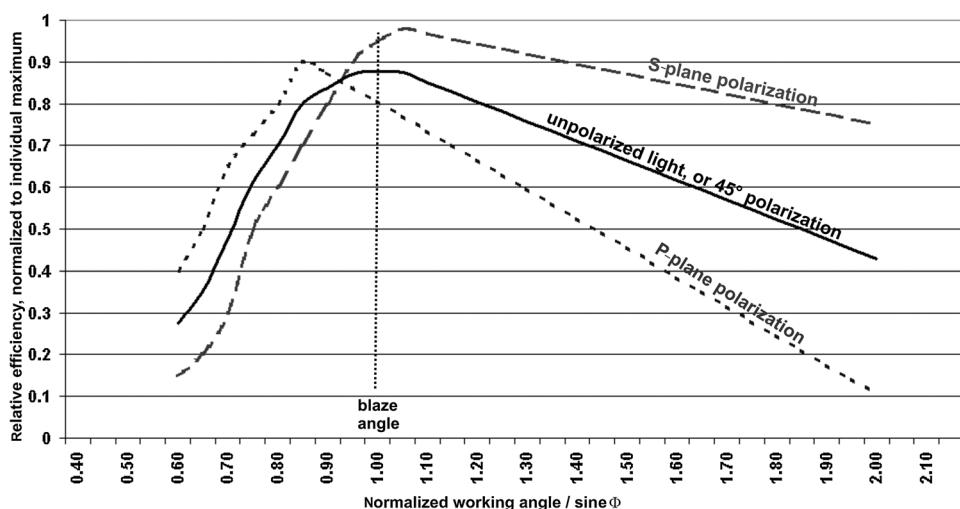


Figure 3.1 General efficiency curves of blazed diffraction gratings.

a grating's useful range within one order is “center wavelength $\times 0.7$ ” for the lower limit up to “center wavelength $\times 2.0$ ” for the upper limit. Both limits represent 50% of the actual maximum efficiency.

3.3 Types of Dispersers

Figure 3.2 illustrates (clockwise from top left) the two versions of holographic gratings, the Echelle grating, the curved grating, and the transmitting version. The standard ruled grating is shown in Fig. 2.2. The following sections go into further detail about each disperser.

3.3.1 Holographic gratings

The good news about this grating is that it can be blazed. The sinusoidal structure can be modified, i.e., by an ion etch process, which will flatten one side of the sine wave. This alteration, combined with the original height of the structures, produces a microreflecting surface very similar to that provided by a ruled grating. Up to 85% of the available light at the actual wavelength can then be dispersed in the “right direction.” It can also be seen why holographic and ruled gratings for the deep UV have a shallow structure: the ratio of the height of the structures to the distance between them produces the angle of the

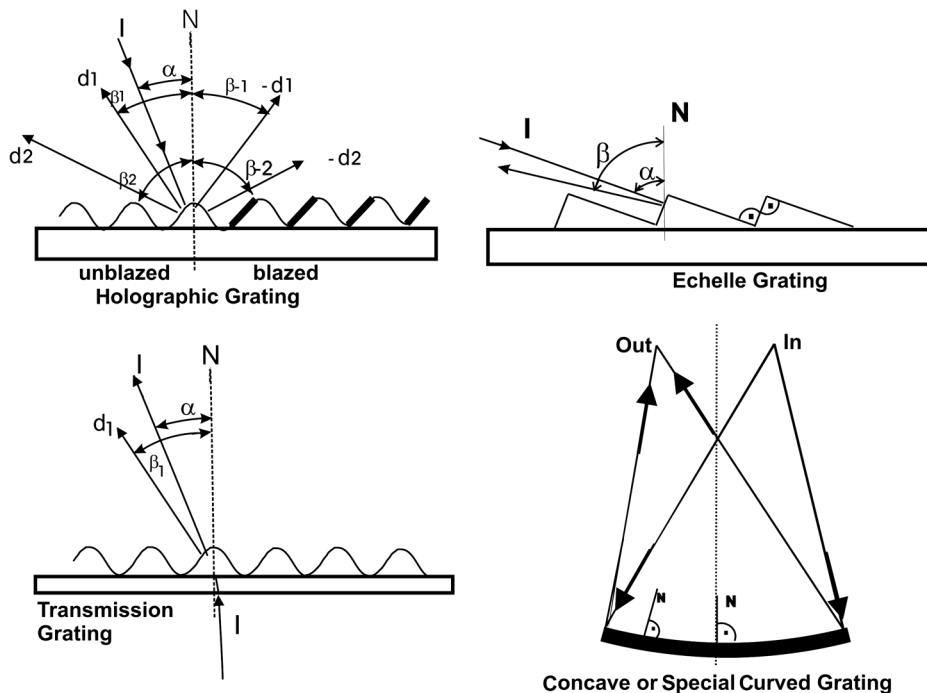


Figure 3.2 Different versions of dispersion gratings.

microreflecting edge, whereas the structures for higher wavelengths and working angles have steeper ruling.

3.3.2 Echelle gratings

A special ruled version is the Echelle grating. In this case, the illuminated side of the sawtooth is the short and steep flank. The microreflection thus appears at high angles, which is good for high dispersion. The corners between the flanks are always at right angles—hence the name (meter, reference). Because of the perfect straight and plane slope, the diffraction has equal, high efficiency over a wide range of spectral orders, which leads to a wide working range covered by a single grating. In fact, Echelles serve the widest spectral range of all gratings; however, also because of the right angles, the grating designers cannot independently vary the angle and length of the illuminated flanks.

Echelles are offered from about $30\text{--}600\text{ mm}^{-1}$ at working angles between 45–80 deg. In use, the angles α and β need to be very close; otherwise, shadowing effects occur, and the microreflection drops quickly. Hybrids between standard ruled and Echelle gratings also exist. If a standard grating with illumination on the long flank works at rather high angles and spreads over many spectral orders, it is called an Echelette grating.

The useful efficiency of an Echelle grating only covers a small angular range. Figure 3.3 represents a grating of 79 mm^{-1} , blazed to 75 deg. Approximately ± 7 deg are useful, but because the grating can be used over a

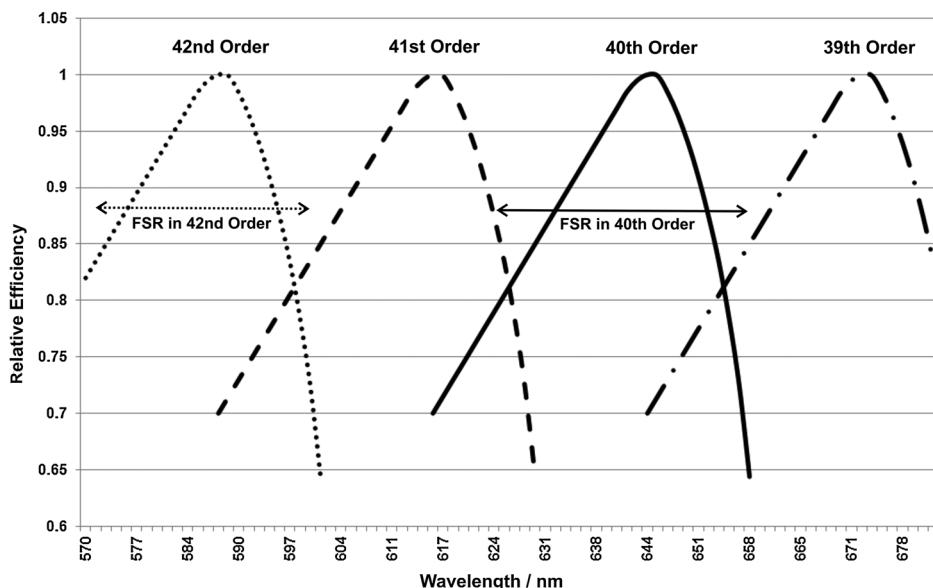


Figure 3.3 Efficiency of an Echelle grating at four subsequent orders.

large number of orders, it serves a range from 180 nm to 24 μm . The example shows the wavelength range of 570–680 nm, which is covered by four orders. The free spectral range is somewhat wider than the useful interval of an order. If the order is changed at the appropriate wavelength, the normalized efficiency can be kept above 80%. Because the diffraction efficiency of an Echelle resides between 70–95% (depending on range and coating), it is a rather efficient disperser. (See Chapter 4 for more information on Echelles.)

3.3.3 Concave and other curved gratings

These gratings combine the functions of a disperser and a mirror. Their advantage is efficiency: one or even two mirrors are saved, which can also lead to a lower total cost and to a smaller spectrometer footprint. Concave gratings are used in compact monochromators and, due to the reduced efficiency of mirrors, in VUV spectrometers. A concave grating is not compatible with flat-field detectors, such as CCD or linear arrays. The focus of the dispersed light only serves a very small area. For spectrograph setups, special curved gratings are used that refocus the dispersed light over a field wide enough to apply multi-channel detectors. These gratings only work in one angular configuration, which means that the spectrograph is fixed.

3.3.4 Transmission gratings

Transmission gratings are very similar to the reflecting gratings. Here, the incident beam **I** first travels through the substrate and then through the structure (or vice versa). The grating rules and formulas remain the same, but the practical application is more complicated. As light moves through the material, the material must transmit within the required spectral range, which is true for both the substrate and the grating structure. In addition, refractive components add to the diffraction function.

The benefit is that it is possible to send the light through the transmission grating in a nearly straight line. This behavior allows for an inline spectrometer, which is desirable if the collimating and refocusing components are lenses. This type of grating is therefore mainly used in fixed setups for the visible or NIR range. If the grating structure is not ruled into the substrate, it comes on film that is bonded to the substrate, and thus firm material and glue need to transmit throughout the applied spectral range (both will create interfaces where the refractive index changes). Because both parts are very thin, the refractive effects might be negligible but appear nonetheless. (Programmable gratings are briefly introduced in Ch. 9.)

3.4 The Prism

A prism is a completely different kind of disperser. Whereas a grating follows the diffraction rules, a prism follows the phenomenon of refraction. Note that a reflecting prism of $A/2$ has a similar but not identical dispersion function

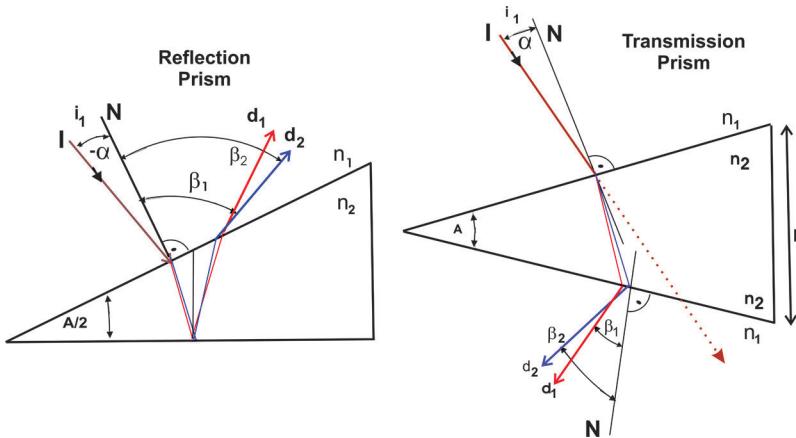


Figure 3.4 The reflecting and the transmitting prism.

compared to a transmitting prism of angle A (see Fig. 3.4). The general refraction formula is the basis of prism calculations:

$$n_{1(\lambda)} \times \sin \alpha = n_{2(\lambda)} \times \sin \beta, \quad (3.1)$$

where n_1 is the refractive index of medium 1 (usually vacuum or air, with $n = 1$), n_2 is the refractive index of medium 2 (the prism material), α and β are the incident and refracted beams, respectively, and N is the normal.

If α and β are on different sides of N , as in the case of gratings, one needs to be marked negative. This fulfills the arithmetic but has no consequence because a prism only produces one output spectrum. An important parameter is the deflection angle d of light leaving the prism, which is defined as the angle between the incident beam α and $\beta_{(\lambda)}$ of interest. For a reflecting prism, the minimum angle d_{\min} is calculated by

$$d_{\min} = \{2 / [\sin(n \times \sin A/2)]\} - A, \quad (3.2)$$

where n is the refractive index of the prism material (n_1 is assumed to be 1), and A is the angle of the prism (the angle of the transmitting version to be used).

The dispersion of a prism spectrometer, similar to the case of a grating, results from angular differences of two wavelengths in the arm of dispersed light. Equation (3.2) calls out the minimum angle d , but in a real spectrometer, the illumination angle i_1 changes as the prism turns, which affects the deflection angle d strongly. Consequently, it also affects the dispersion. The deflection angle d in a spectrometer can be calculated by

$$d = i_1 + \frac{1}{\sin \left[(\sin A) \times \sqrt{(n^2 - \sin^2 i_1)} - (\sin i_1 \times \cos A) \right]} - A, \quad (3.3)$$

where d is the deflection angle, A is the angle of the prism (in reflection mode, $2 \times A/2$), n is the refractive index of the prism material at the wavelength

considered, and i_1 is the incidence angle of the illumination related to the normal N .

The angular dispersion can be calculated by chained application:

$$\delta d = d(\lambda_2) - d(\lambda_1). \quad (3.4)$$

Similar to the grating, the spectrometric reciprocal dispersion RD is found by

$$RD = 1/[f \times (\sin \delta d/d\lambda)] = d\lambda/(f \times \sin d\delta), \quad (3.5)$$

which includes the focal length f .

The resolution results from the dimension of the base length b and the refractive index at the wavelength considered:

$$\lambda/\Delta\lambda = -b[dn/d\lambda], \quad (3.6)$$

where b represents the base length of the prism. For example, if a prism's $n = 1.6$ at the λ of interest, and the angle $A = 30$ deg, d will be 15 deg.

By selecting a useful material, prisms are available from the VUV to the far-IR. Compared to gratings, prisms have the advantages of order-free dispersion, a wide useful range, high transmission, and lower stray light. Disadvantages are the change in dispersion over wavelength, the dependence of the output angle on changes of the incident angle, and the travel of the reflectance angle. They require a more complicated drive control than gratings. However, it can be useful to combine a grating and prism, as described in the following section. (The application of prisms is described in Section 4.1.8.)

3.5 The Grism

A rather exotic disperser, the grism consists of a grating that is attached directly to a prism. This arrangement has the advantage of sorting the upper spectral orders and providing higher total dispersion. There are two basic versions of grisms: the transmitting type, where the light travels first through a grating and then through a prism, or vice versa; and the reflecting type, where the prism is illuminated first, and the grating is attached to the rear side. In the latter case, the prism disperses twice. A third version would involve a transmitting grating in front of a reflecting prism, in which case both elements are passed twice, which provides enhanced dispersion.

In any case, the angular and dispersion behavior is complex because it is the result of a chain of at least two different processes. In the transmission version, the grating structure and the prism need to be fully transparent in the required working range. Grisms are often masters, with the grating ruled into the prism material. In the past, before microprocessors were available, the wavelength control was too complicated for serialized spectrometers. However, that limitation does not exist anymore, and the routine application of grisms is just a matter of will. Grisms are actually applied in special applications that either require minimal space, weight, or number of components. In satellite or military applications, all three factors apply.

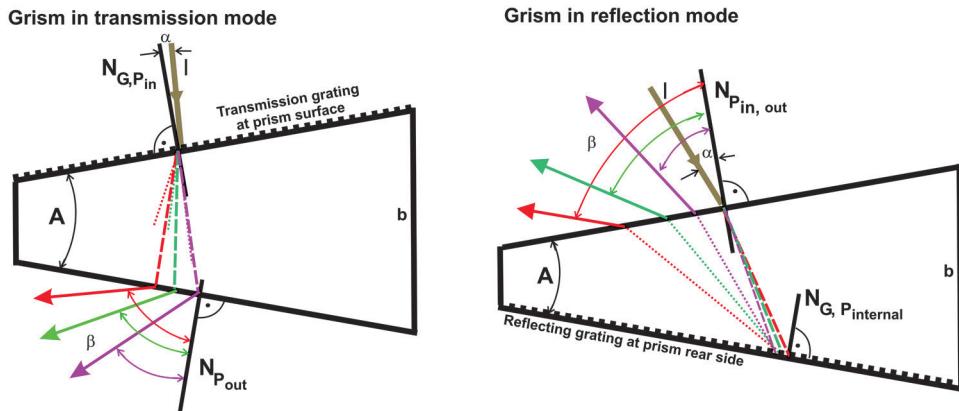


Figure 3.5 Two basic grism versions. N_P is the prism normal, and N_G is that of the grating. The other signs are identical to grating and prism equations. The figure illustrates Littrow prisms featuring a cut angle A .

Figure 3.5 features two of several possible configurations. The angles drawn represent real interrelations. On the left is a transmission model consisting of a grating with $k = 1.67 \mu\text{m}$ (600 mm^{-1}), illuminated under $\alpha = 5 \text{ deg}$. The prism is made of quartz, with angle $A = 15 \text{ deg}$. The dotted lines after the grating display the dispersion of the grating alone, whereas the dashed lines incorporate the refraction of the prism. It is assumed that the grating is either ruled into the prism or else the substrate and bonding do not modify the ray travel. The rays will leave the grism under 27 deg at 250 nm , 41 deg at 500 nm , and 55 deg at 800 nm . That is a mean dispersion of $17.9 \text{ nm}/\text{deg}$ in the UV and $21.5 \text{ nm}/\text{deg}$ in the red wavelength range, which is more homogenous than the variation of dispersion by a grating or a quartz prism alone. The reflecting version on the right presents the results of a similar element. The grating is reflecting and also acts as the rear side of the prism, which can be created by ruling the grating into the prism surface and coating it with a reflecting material. Again, the angles represent reality, $\alpha = 15 \text{ deg}$, in order to enhance the first prism dispersion. Because the prism is passed twice, it dominates dispersion, and the typical increase in angles towards the UV takes place. The data, related to N , are 22 deg at 250 nm , 49 deg at 500 nm , and 70 deg at 800 nm . The mean dispersion is $9.2 \text{ nm}/\text{deg}$ in the UV and $14.2 \text{ nm}/\text{deg}$ in the red wavelength range.

Grisms are calculated by chained equations, allowing for a wide variety of combinations. This also allows for the production of an element that either disperses “straight through” or “straight back.” Both mean that the rays travel under the same mean angle as the illumination beam I, while the dispersed light travels along both sides of I.

3.6 Other Features of Diffraction Gratings

3.6.1 Polarization anomaly

The source of the polarization anomaly, also called “Wood’s anomaly,” is the interaction of light with the surface structures, creating interference patterns that overlay the constructive diffraction patterns. If the actual wavelength is close to k or a full multiple or division of k , Wood’s anomaly can affect the signal. The anomaly almost only affects the S plane and not much of the P plane.

In Fig. 3.6, the line frequency of the ruled replica grating is 1200 mm^{-1} , at a blaze angle of 17 deg, or 500 nm under Littrow condition. When all P-plane light is fairly homogeneous, the S plane shows strong local interference. It also reveals that Wood’s anomaly does not need to be a loss of signal; it can also produce overshot. The resulting curve for “nonpolarized” light shows that the addition of P and S reduces the anomaly’s impact but does not mend it. If one is working with a system like this and the experimental polarization ratio changes between measurements, there will be a problem with signal comparison. In general, anomalies are stronger with holographic gratings compared to ruled ones. Echelle gratings suffer relatively little from Wood’s anomaly.

The possible strength of Wood’s anomaly is the topic of Fig. 3.7, which presents some measured examples with gratings of 1200 mm^{-1} . All curves are normalized to the same scale, and all have been recorded with polarization-scrambled light. Curve 1 (lamp emission) is the physical behavior of the object, a tungsten-halogen lamp. UV–Vis data acquisition (curve 2) was performed with a phototube (see Section 5.2.1); curves 3 and 4 used a UV–silicon detector (see Section 5.6). All conditions were kept constant. Three gratings of 1200 mm^{-1} ,

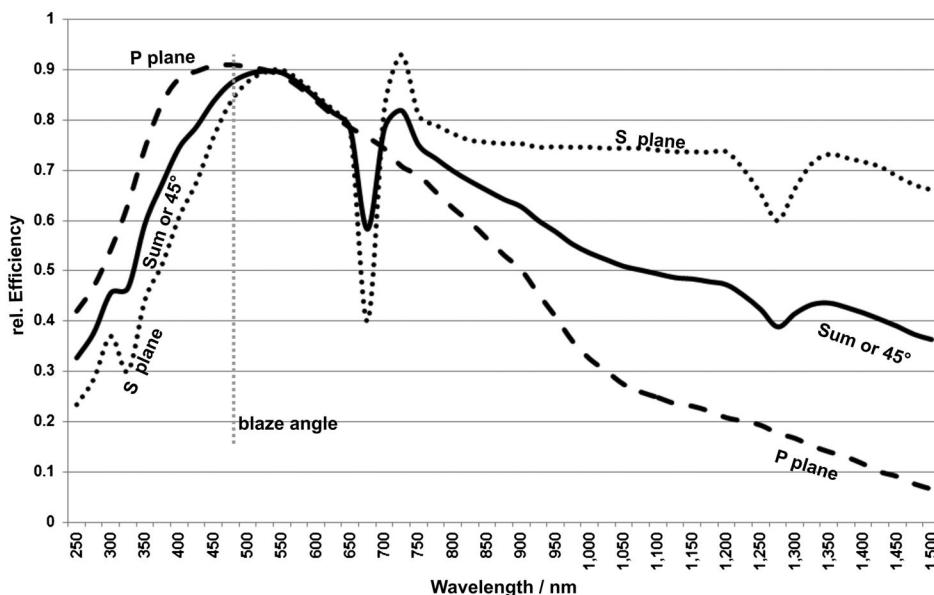


Figure 3.6 Example of polarization anomalies.

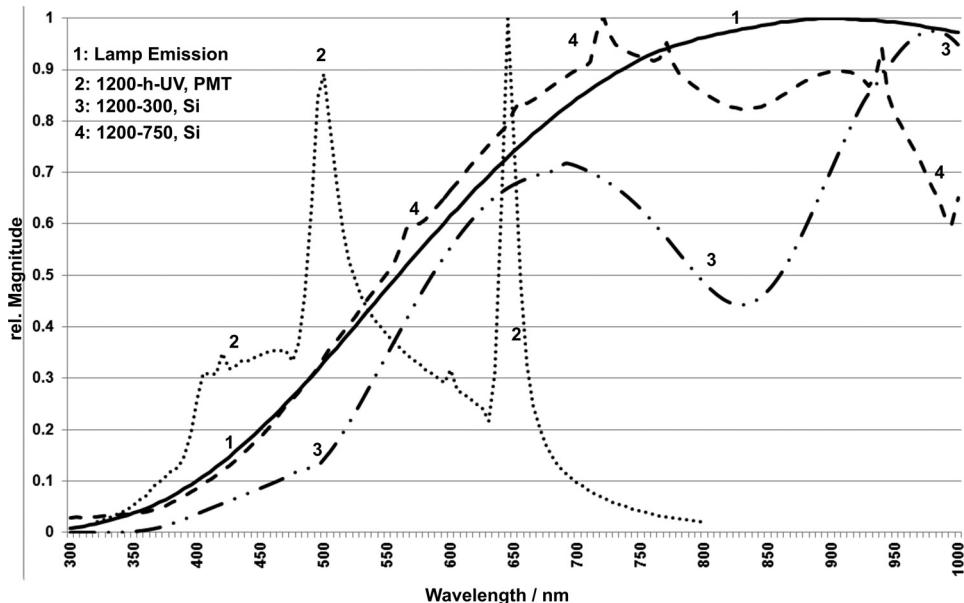


Figure 3.7 Examples of how Wood's anomaly can affect the measured result.

but different blaze, were examined on a rotating table. The first one, curve 2, was a UV holographic 1200 mm^{-1} , with 8-deg blaze or 250 nm. The two maxima at 500 and 650 nm are constructive Wood's anomalies. Curve 3 represents a grating of 10-deg blaze (ruled, 300-nm blaze). The curve is rather continuous but shows wide suppression between 700–950 nm. Curve 4 represents a grating of 26-deg blaze (ruled, 750-nm blaze). In this case, a slight dip is also present, superimposed by some small, constructive interference peaks. The data prove that Wood's anomaly does not only depend on the line frequency.

3.6.2 Polarization of Echelle gratings

Figure 3.8 shows the total efficiency and both planes of polarization of the Echelle grating (presented in Fig. 3.3) for two spectral orders. The S plane is preferred when the P plane first follows the S plane, but already below the blaze angle, it starts to decay. Both planes reveal rather few anomalies. The general efficiency and polarization is similar for all Echelle gratings, whereas Wood's anomaly changes with the model and spectral order.

Why do Echelles show less variation than gratings of higher density? Anomalies are effects of waves and matter. A grating of 1200 mm^{-1} will experience standing waves at the corresponding wavelength with the center at 833 nm. Multiples of that wavelength can also be affected. In addition, the gratings have no blank surface—they are almost always coated, and interference can follow. Echelles, on the other hand, are rather coarse structures; line densities above 632 mm^{-1} are not available. The right-angle structures [see Fig. 3.2] provide additional support for a smooth polarization response.

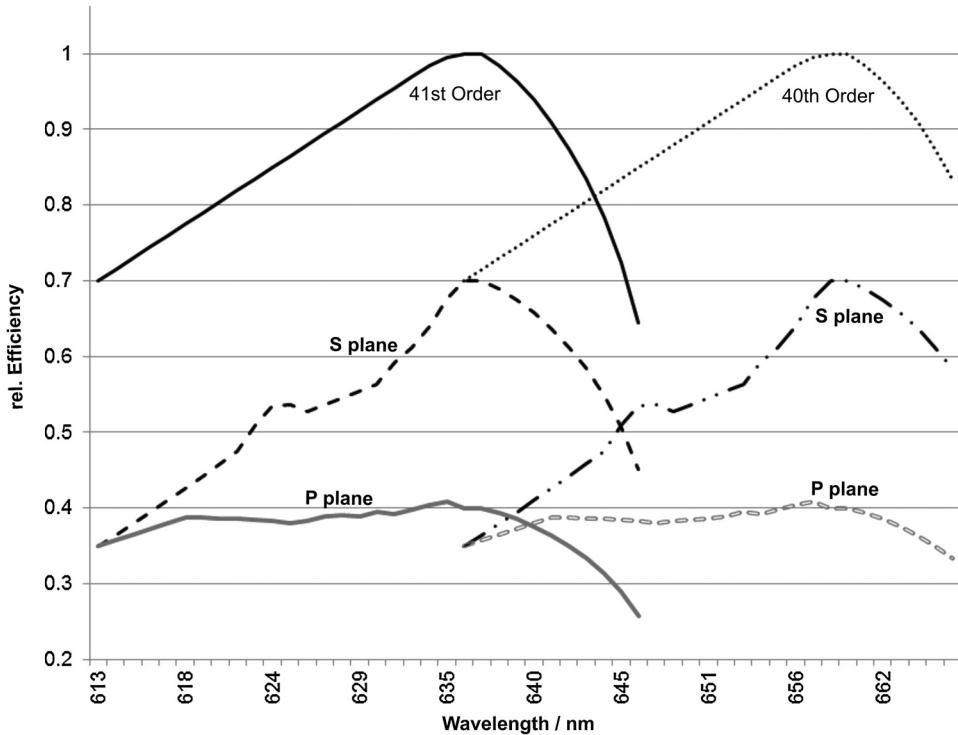


Figure 3.8 Polarization and anomalies of Echelle gratings.

3.6.3 Scattering effects

A ruled grating was made by changing the surface structure through the application of a diamond or a master grating. Unfortunately, processes like that will rarely create clean edges. In most cases, the crest of a line structure will be imperfect (the same is true for the valley). The irregularities scatter some of the incident light. It is not possible to give general rules for the magnitude of scatter, but due to the fabrication process, a holographic grating almost always scatters significantly less compared to a similar ruled one. The ratio can reach up to a factor of 10. The minimum scatter comes from holographic gratings without additional treatment, which means “unblazed” sinusoidal ones. Scattering effects can be reduced by applying a coating process right after production. The coating will average out the production residuals and preserve the surface from oxidation.

3.6.4 Grating ghosts

A good holographic grating will have precisely the same distance k between all line structures of the surface. A ruled grating is done on a machine with inherent tolerances; however, even an interferometric controlled-ruling engine can change the k slightly from line to line. It is also possible that the diamond

slightly moves sideways while ruling one line. It is valid for both versions, ruled and holographic, that the substrate surface might not be perfectly flat. It can have some faint variations. For the light arriving under an angle, that will also look like a small change in k over the surface. For example, consider and estimate of a variation of zero and -0.5 nm . If the grating is nominally 1000 mm^{-1} , $k = 1000\text{ nm}$. For an α of 20 deg , the expected first-order wavelength is 684.041 nm . At $k = 999.5\text{ nm}$, it is 684.383 nm , providing a difference of 0.342 nm .

There are many possible spectral distortion patterns, of which two are considered here:

1. If k statistically varies between 1000 and 999.5 nm , there will be no possible resolution better than 0.684 nm —even sharp peaks will be displayed with a flat or wide top!
2. If 90% of all lines are at a distance of 1000 nm , but the other 10% are at 999.5 nm , there are two peaks 0.342 nm apart and at an amplitude ratio of $90:10$.

This behavior is where the name “ghost” comes from. If an experimenter finds a peak “where there is no peak,” it might be due to a ghost. There are two ways to get rid of ghosts: replace the grating, or use masking if the grating has distortion only in a limited peripheral surface area.

3.6.5 Shadowing and diffusion

If the working angle supersedes the range planned for a grating, or if the angle between α and β is large, the sawtooth structure can shadow part of the active flank. Not only does the total efficiency drop, but the dispersed light, traveling from the flank towards the output mirror, can interact with the crest of the neighbor flank and be diffracted there again. This behavior diffuses the light beam, which affects both the resolution and the contrast. A different kind of diffusion would be evident if light is not reflected and diffracted in the upper layer of the surface. If it penetrates the surface before being reflected, refraction is added to the diffraction, and the increased absorption reduces the reflection. The wavefront will be distorted, which also leads to a loss in efficiency, contrast, and resolution. This could be the case if a grating with an aluminum layer was not properly coated and thus aluminum oxide, which is semi-transparent, has built up.

3.6.6 Surface coating

Quality gratings either have an inert surface or come with a coating. The standard structure material, aluminum, has a reflection coefficient around 0.89 through the UV–IR range. For the UV–Vis–NIR range, the most often used coating is a mix of Al–MgF₂ (sometimes with quartz). It improves the average reflection efficiency of aluminum, protects the surface, and prevents oxidation. Other coatings, such as iridium, gold, and silver, will increase the efficiency in a

certain wavelength range: up to 96% reflectance efficiency is possible. Because the electromagnetic parameters of the modified layer material and thickness are different and the structure is slightly changed, the blaze, polarization behavior, and anomalies can be modified, too.

References

1. J. M. Lerner and A. Thevenon, *The Optics of Spectroscopy*, Horiba Scientific website, <http://www.horiba.com/us/en/scientific/products/optics-tutorial>.
2. C. Palmer and E. Loewen, *Diffraction Grating Handbook*, 5th ed., Thermo RGL, Rochester, NY (2002).

Chapter 4

Design Considerations of Monochromator and Spectrograph Systems

This chapter extends the topics discussed in Chapter 2 and describes real-world parameters that need consideration when defining, designing, or selecting a spectrometer. The parameters and equations presented are very helpful for the design, selection, and calculation of a spectrometer and the components used. With regard to construction, it is recommended to apply ray tracing before building and testing a prototype. Even though ray-tracing programs are rather complicated to use, they will show all pathways of light travel. Some programs even provide guidelines to help improve the setup. Complicated designs, such as those required for imaging corrected spectrographs, require ray tracing. In terms of construction considerations, no light is to be shed on optical holders or other mechanical parts. It is assumed that the information available from suppliers is sufficient to select stable and useful kinematic holders as well as other mechanical components. (Mechanical stability and thermal impacts are discussed in Section 2.7.) Compact, integrated spectrometers are not discussed because they offer no access or flexibility to the customer—they either fit or they don't.

4.1 Beam Travel inside a Spectrometer

Figures 4.1 and 4.2 show the general beam travel in a spectrometer featuring a Czerny–Turner (the rays travel similar to an Ebert–Fastie) with a symmetric grating inclusion of $\delta = 30$ deg and a grating that turns towards either the entrance or the exit, respectively. One difference is an inversion in the spectral orientation, which has two consequences:

1. the orientation at the surface of an area detector flips, and
2. in multiple-stage spectrometers, the additive and subtractive function can be switched.

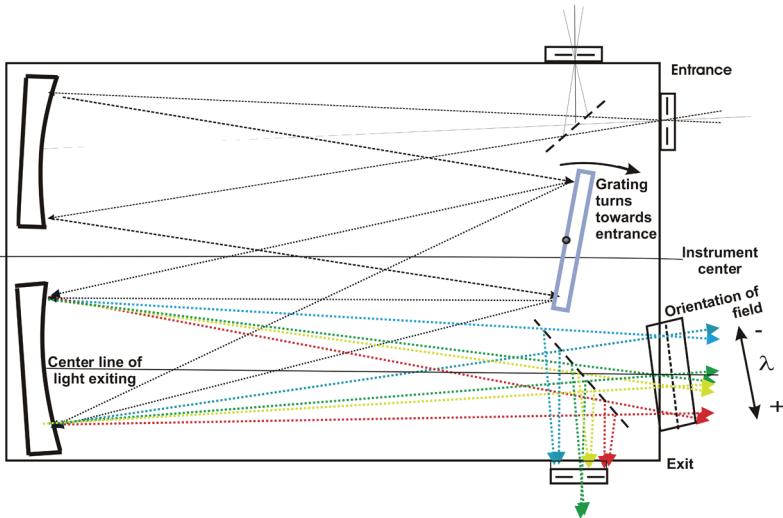


Figure 4.1 Beam travel within a Czerny–Turner where the grating turns towards the entrance.

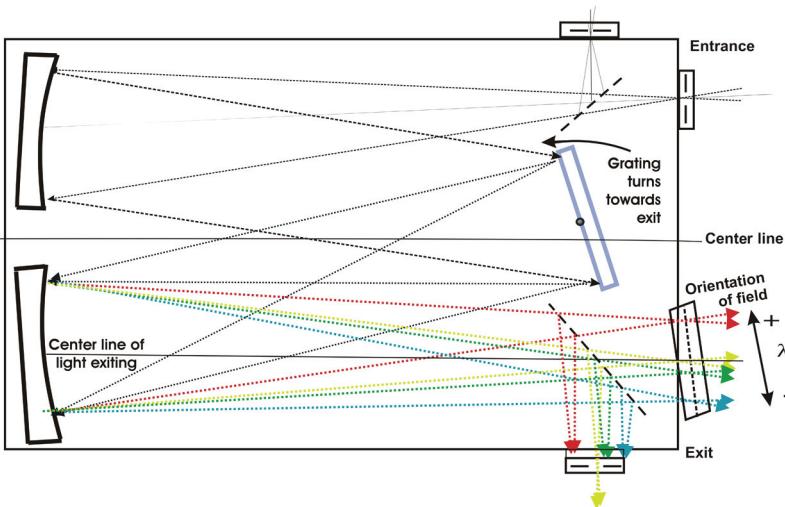


Figure 4.2 Beam travel within a Czerny–Turner where the grating turns towards the exit.

In principle, it matters little which direction the grating turns—both directions have benefits and drawbacks. Consider an example based on a spectrometer with an inclusion angle $\delta = 30$ deg. The mirrors are considered to show some reserve in width, 110% of the grating width. Obviously, the projected height of the beam is constant, whereas the width will change with grating rotation (see Sections 2.3.1–2.3.3). Under the reflection condition (zero order), the grating is angled 15 deg to both mirrors, and 3.5% of the projected

area is lost in each direction. If the grating turns towards the entrance, it reaches full projection at $\alpha = 0$ deg and $\beta = 30$ deg. This seems very useful because the dispersed light leaves the grating fan-shaped and still fills the focusing mirror; thus, the output aperture is fully used. On the other hand, the light can now travel back and forth between the collimator and grating, producing multiple dispersion, which can be disruptive if it overlays unwanted wavelengths. As soon as the range of that reflection is left behind, more light is available compared to the alternative grating rotation, which turns to the exit.

This method reaches full area projection between the grating and focusing mirror, when $\alpha = 30$ deg, but at that point the grating has already lost 13.5% area efficiency, as seen from the entrance. Those losses will increase as the grating turns further, but internal reflection and multiple dispersion will be considerably less: only discrete wavelengths can be involved, but not the full spectrum, as occurs when turning towards the entrance.

Spectrometers are available that provide correct grating rotation in both directions. If that option is utilized, the orientation of dispersion will flip, and the blaze behavior of the gratings will probably change significantly. It is good practice to check both directions with a known source and then compare the results.

4.1.1 Beam travel in a symmetric spectrometer

The examples shown in Figs. 4.1 and 4.2 each have two inputs and output ports: two entrances and the diverting mirror, plus two exits, one with a slit and the other for the illumination of a field. To support a wide output field, the grating has been moved slightly out of the center of the system, as marked by the grating's center dot and the center line. Thus, the front entrance beam enters under an angle not perpendicular to the front wall, whereas the side-entrance diverting mirror compensates for that angle. In the example shown, the dispersed, exiting light travels parallel to the side wall, and the plane of the front exit field is also tilted versus the wall. The angles result from construction specifics and are typical for a certain series of spectrometers. In addition, the field's tilt angle might be different for the best spectral focus distribution, compared to the best imaging focus behavior. The focal function is not influenced by the grating rotation or whether the grating is placed in the center of the spectrometer. Whether the grating turns around the front axis or not will not affect the focus behavior. The slight asymmetry was chosen in favor of a large nonvignetted output field while keeping the internal angles as narrow as possible. The sketches so far do not show that the used surface of the focusing mirror, and consequently the output distribution, change with the wavelength and grating angle.

The outer regions of area detectors do not receive the same amount of light as the center. A slit output does not suffer from that deficit. In Fig. 4.3, only those wavelength rays traveling inside the yellow range have optimal transfer. The outer wavelengths only partly hit the focusing mirror. If the mirror has the same width as the grating, the grating's working angle decides

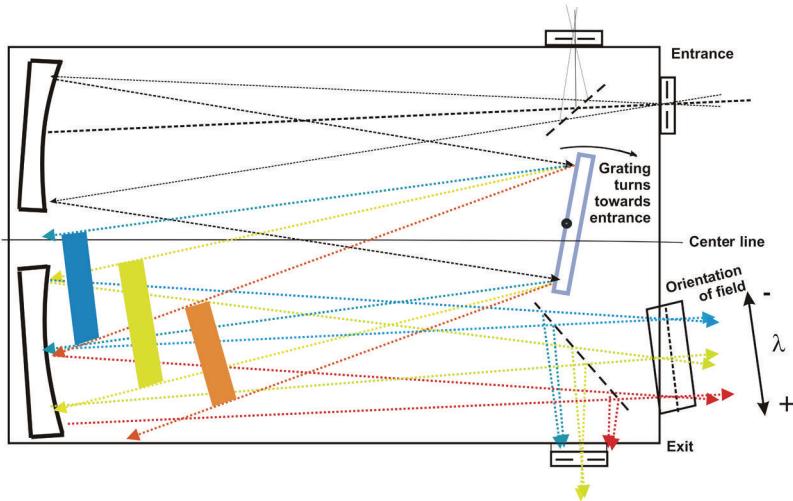


Figure 4.3 Restricted area of full illumination.

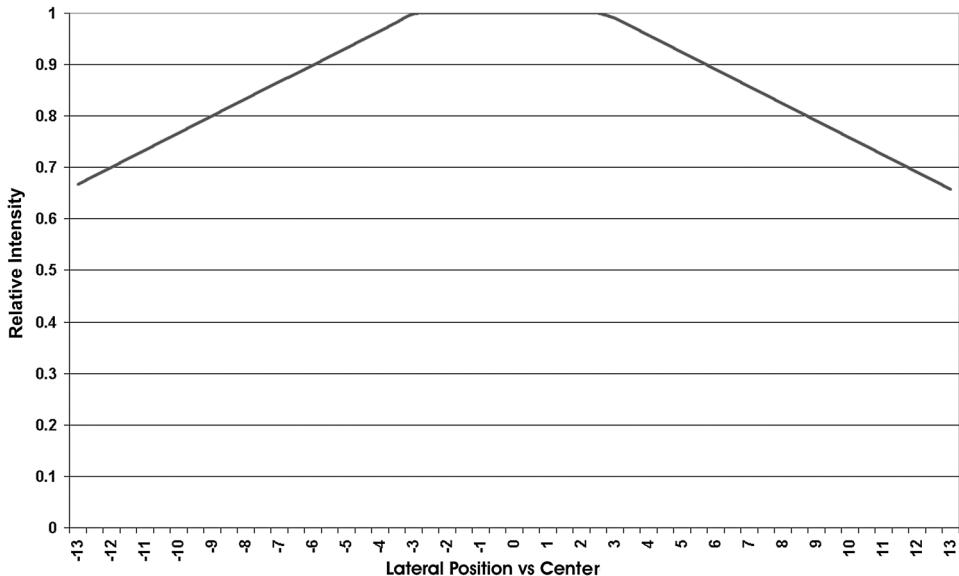


Figure 4.4 The typical intensity distribution over the horizontal width of the field.

the actual used width. In principle, the output field area receives an illumination like that shown in Fig. 4.4.

The extension of the optimum plateau, the shape of the curve, and the steepness of the edges depend on the actually used grating width W , the grating's angle ($\cos \alpha$ or $\cos \beta$), effective width of the focusing mirror, shadow effects at the grating, blaze/transfer behavior of the grating, and homogeneity of illumination. The curve can take different shapes, and the effect is called "vignetting."

If the focusing mirror is oversized, the width of the plateau increases at the expense of the resolution and imaging quality, as discussed in Sections 2.4 and 2.6.6.5. A compromise must be found. One more disadvantage of increased width is the increase of internal angles, which affects the Q factors and complicates imaging correction. The *f*-number will not increase, but the resolution will drop and stray light will increase due to the larger surface.

Consider an example with the following specifications: $f = 500$ mm, $\epsilon = 8$ deg, detector length = 25 mm, center wavelength = 580 nm, grating width = 80 mm, width of focusing mirror = 85 mm, and groove frequency = 600 mm^{-1} . That will lead to a dispersion of about 3 nm/mm or a 75-nm interval at the detector. The colors in Fig. 4.3 fit those wavelengths. The grating rotates towards the entrance. At 500 nm, the working angle will be 9.92 deg, leading to a surface factor of 0.985 and a used width of 78.8 mm. These dimensions will make ± 3.1 mm of fully illuminated side wings beside the center of the mirror. Translated to the 25-mm detector width, there will be an additional ± 0.91 mm at the flat top. In Fig. 4.4, the flat-top range increases from 3 mm to 4 mm in both directions.

4.1.2 Variations of the basic Ebert–Fastie and Czerny–Turner concepts

Many variations have been realized, and some leave enough space for large fields beside the grating, thus still offering symmetry:

- The front entrance and exit mounted oblique to the wall, as shown in Fig. 4.5. The “W” configuration has the advantage of providing full symmetry. The out-of-plane angles can be made identical, as well as the

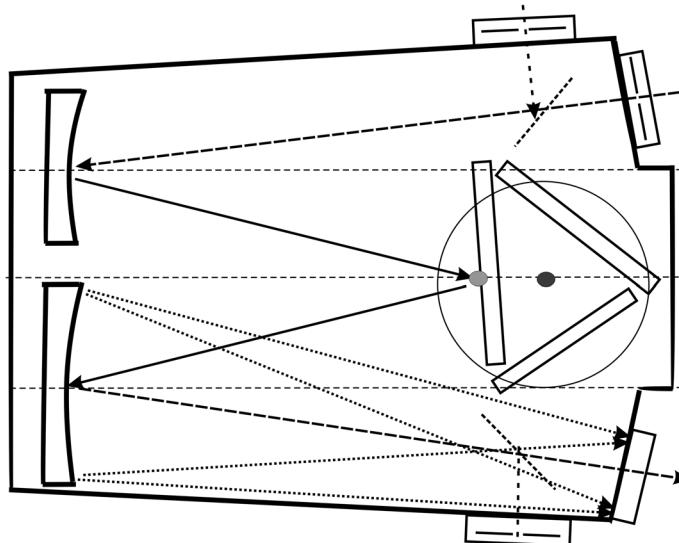


Figure 4.5 General beam travel in an oblique Czerny–Turner spectrometer, also called the W configuration.

two ε . The advantage can easily be derived from the known equations. Even an enlarged focusing mirror, as shown in the figure, keeps symmetry. A grating table for up to three gratings was used for the image, which leads to a general discussion of multiple-grating constructions.

- Single-grating versus rotating configuration. Consider a single grating installed as shown in Figs. 4.1–4.3: The pivoting point will remain at the center of the grating surface, and the dispersing element will always rotate in the plane of its front surface. In Fig. 4.5, the whole table rotates around the black dot, while the active grating swings through the illumination beam. Therefore, the grating's center (grey dot) will move while the grating change angles. The implications of this setup are discussed in Section 4.2.2.
- Crossed beams inside the spectrometer housing. This arrangement produces shallower angles at the mirrors and the grating as well, as shown in Fig. 4.6. A similar Czerny–Turner has also been developed. The construction “extends” the focal length and is less susceptible to stray light than the “parallel” design. The internal angles avoid the direct transfer of light from the entrance to the output. The entrance and exit can only be placed at the side, which limits flexibility.

4.1.3 Output wavelength as a function of the source position

What happens if the object's emission (light source, lens, mirror) does not enter perfectly at the axis of the spectrometer? Or if the input light does not travel perfectly parallel to the axis? In many applications, users are clever enough not to use the full aperture of the system, leaving some reserve for grating movement and other changes. This way, a small movement of the

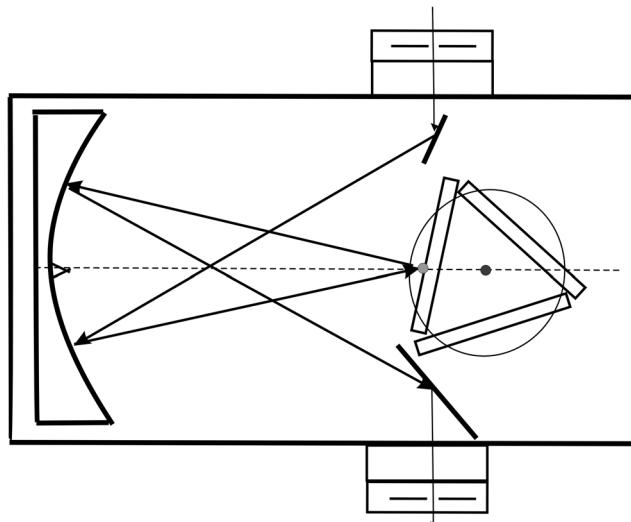


Figure 4.6 Crossed Ebert–Fastie design.

source away from the optimum angle and position will not change the output amplitude. An 8-mm sideways variation at the collimator of a 500-mm spectrometer only effects an angular displacement of ~ 1 deg. But even if the amplitude stays the same, the impact on the wavelength at the output center can be calculated by Eq. (2.1). A comparison of gratings of (a) 300 mm^{-1} and (b) 1200 mm^{-1} with a laser wavelength of 632.8 nm demonstrates the impact:

- a. A spectrometer with $\varepsilon_1 = \varepsilon_2 = 8$ deg will require $\sin \alpha = -0.0436$ and $\sin \beta = 0.23345$ at the grating. α is 2.5 deg and appears negative because the beam travels on both sides of the grating normal N ; β is 13.5 deg. Consider that a sideways shift in illumination of only one-fourth of a degree produces $\alpha = 2.25$ deg and $\sin \alpha = -0.0392$. Because angle β stays constant, the output wavelength jumps to 647.5 nm, a difference of +14.7 nm. In a 500-mm instrument, the dispersion is $\sim 6 \text{ nm/mm}$; thus, the geometric jump is >2 mm in the output field.
- b. The same experiment with a grating of 1200 mm^{-1} , where $\sin \alpha = 0.21225$, or 14.55 deg, and $\sin \beta = 0.5083$, or 30.55 deg. Both angles are at the same side of N . A source displacement by one-fourth of a degree produces $\alpha = 14.3$ deg or $\sin \alpha = 0.247$. The resulting output wavelength is now 629.4 nm, a change of -3.4 nm. With the 1200 mm^{-1} grating, the dispersion is $\sim 1.5 \text{ nm/mm}$, again creating a geometric move in the output field of >2 mm. The direction has changed because α and β are now on the same side of N .

In short, the axis of illumination plays an important role during wavelength calibration and for precise results.

4.1.4 Local output dispersion as a function of the lateral position in the field output

Figure 4.7 demonstrates how the dispersion changes over the lateral axis of the field output. The reciprocal dispersion (RD) depends on the sine functions of the grating's working angle and on the focal length. The dispersion can be calculated with the help of equations such as Eqs. (2.17)–(2.20). The most convenient is Eq. (2.20): $RD = \lambda / (2f \times \tan \phi)$. The working angle ϕ results from α and β ; at least one of the two varies with wavelength whether the grating moves or not. Because $\sin \alpha$, $\sin \beta$, or both vary differently than $\tan \phi$, the relation between λ and $\tan \phi$ is different at any spectral point and also over the spectral position in the field output. A nonlinear function occurs that requires flexible software efforts in the calibration routines of spectrograph/array systems.¹ The graph represents the behavior that appears if the grating rotates towards the entrance to produce increasing wavelengths. Angle β is always smaller at the lower-wavelength end of the field, leading to a higher RD number compared to that at the center. At higher wavelengths, the angle β increases, leading to a lower RD value (the wavelength steps are more expanded/dispersed).

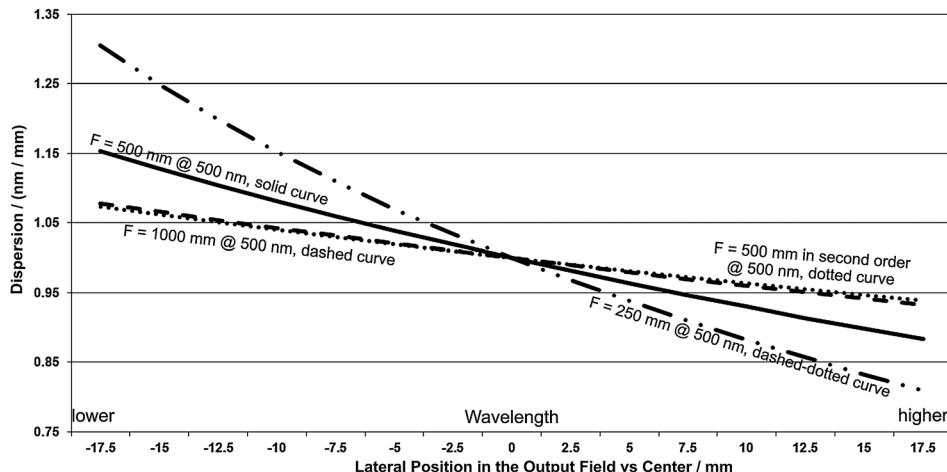


Figure 4.7 Progression of four typical dispersion curves, distributed over the lateral output field of a spectrograph that is equipped with three different gratings turning towards the entrance.

Figure 4.7 shows three spectrometers with basically the same grating and mirrors, but different focal lengths. It is assumed that the grating is 1200 mm^{-1} and works at $\phi = 17 \text{ deg}$. The mirrors are 80-mm wide, and the output field is 35-mm long. An instrument with $f = 250 \text{ mm}$ (dashed-dotted curve) will have an f -ratio of 3.125 and rather wide internal angles that create an RD variation between $+30\%$ and -20% . The $f = 500 \text{ mm}$ instrument provides $f/6.25$ at remarkably narrow angles and little variation in RD, whereas the $f = 1000 \text{ mm}$ instrument at $f/12.5$ provides almost-parallel ray paths, leading to a very small change in RD over the detector width. It is interesting to note that the variation will grow for gratings for small working angles (little groove frequency), whereas the relative variation decreases for gratings with high groove density. If large wavelength intervals are guided to the detector, they will have a rather large change in dispersion, which is proven by the fourth curve: the $f = 500 \text{ mm}$ system again but with a center output wavelength of 1000 nm (dotted curve). Because the sine value of the grating angle in second order is twice that at 500 nm, the relative variations are considerably less. It also proves that an $f = 500 \text{ mm}$ instrument in second order provides the same RD as an $f = 1000 \text{ mm}$ system in first order if all components and angles are comparable. If the grating would rotate towards the exit, the curves would be inverted, with the lower value on the left of the graph and the higher nm/mm numbers on the right.

4.1.5 Output dispersion and fidelity as a function of the tilt angle of the field output

The angle of the output focal plane in relation to the housing and/or the center axis of the traveling light might be tilted. This angle is sometimes called the

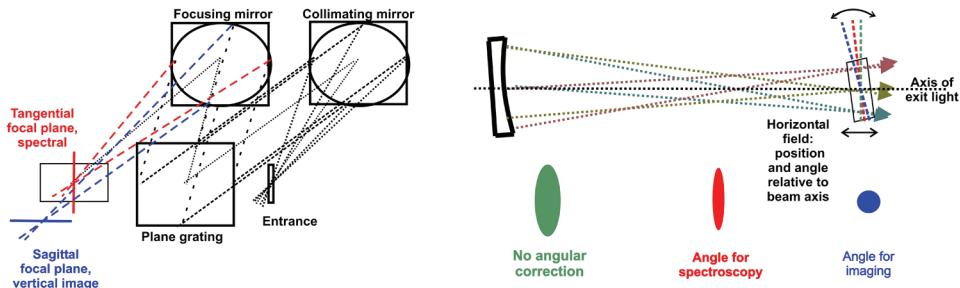


Figure 4.8 Typical signal shapes at a spectrometer output.

“tilt angle” or “detector angle;” it is generated by the internal angles of the system, which influences the position of the foci at different lateral positions. Because the foci need to be correlated with the vertical distribution, as well, there might be a different angle for the horizontally oriented rays compared to the vertically oriented. If that is the case, there will be different angles for the best imaging and spectral focus. Some of the figures in this chapter account for the scenario where the output plane might be tilted.

The left side of Fig. 4.8 illustrates beam travel in a Czerny–Turner. If all rays enter the spectrometer through the same spot, then they have slightly different pathlengths to the collimator due to the tilt angle of the mirror. As the grating turns, the parallel rays from the collimator must travel different distances. In summary, the rays traveling in the horizontal plane have a shorter path compared to those moving vertically. The horizontally oriented rays, which have the same orientation as the wavelength dispersion, are called “tangential;” they are responsible for the spectral focus. The vertical rays are mainly responsible for the reproduction of images or vertically separated spectra; they are called “sagittal,” and they refocus at different places. The effects can be partially compensated for by tilting the output field. If the plane is not perpendicular to the axis, the detector needs to be tilted in the spectral direction. If that is the case, and the angle is not corrected but perpendicular to the beam center, poor reproduction occurs. If the tangential- and sagittal-oriented rays have pathways of equal length, the optimal foci for both can be reached.

The right part of Fig. 4.8 demonstrates three situations. To reach the best spectral resolution at the expense of the vertical resolution (and vice versa) might require different tilt angles. Some types of spectrometers are equipped with variable output flanges to allow for easy angle changes. Even if no optimum is found for both applications, there is a compromise position between the two best positions. The shorter the focal length and the wider the internal angles of the spectrometer, the more the tilt angle can differ from perpendicular and the larger the angular difference between the “two best” positions can be. Tilt angles of up to 35 deg are known. The impact on the spectral and the image position in the output is a cosine function of the tilt

angle. Dispersion and spectral interval, arriving at the detector's surface, are influenced linearly. They must be entered into the calculations. Because tilting the output plane will most often not be sufficient to correct the aberrations created by the use of entrance spots other than the center of the slit, more corrections are required for multi-stripe spectroscopy or imaging.

4.1.6 Correction methods for spectral Imaging

In order to align the traveling pathways of tangential- and sagittal-oriented rays, which do not travel through the vertical center of the spectrometer, more optical corrections are often required.² They depend on the internal angles and the size of the illuminated output field. Four methods have been developed. The design of the correcting optics in all cases is rather complicated and requires ray-tracing software. The applied correction optics are always system specific and are only useful in the pre-defined setup under reproducible conditions.

4.1.6.1 External imaging correction

Based on the design shown in Fig. 4.9, it might look complicated to deform the beam in front of the main spectrometer, but it has some advantages. The third mirror of the entrance attachment (the shaded “Correcting Mirror”) in the figure has a curvature that compensates for the differences in pathlength inside the spectrometer and the attachment itself. The shape of the curvature needs to be found by ray tracing.

The attachment allows conversion of the experimental cone of light to the cone required by the spectrometer. The “Collimator” performs that task and guides the light further in a collimated manner, where the option for more optical components, such as a filter, is provided. After the correcting mirror, the “Focusing Mirror” illuminates the entrance slit by the required angle. The entrance of the attachment can be a physical slit, a variable aperture, or the end of a fiber cable (probably with a number of fibers in a vertical row).

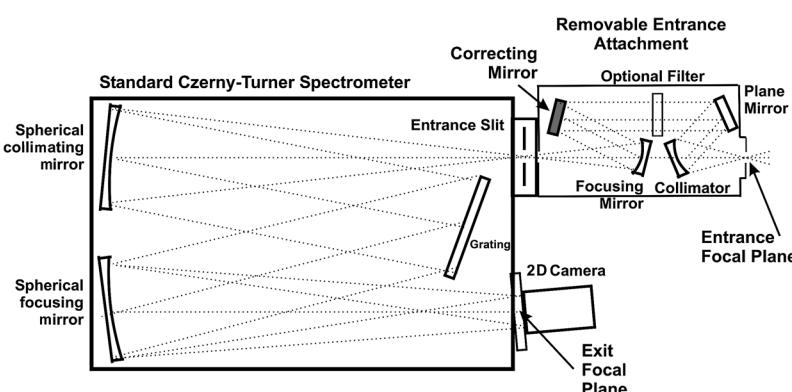


Figure 4.9 An imaging spectrometer with external optical correction.

At the spectrometer entrance the image will appear deformed, but at the detector it will be correct.

Advantages

- The attachment can be taken away and re-installed easily,
- It includes an *f*-number converter,
- It reduces stray light,
- It provides additional space and can be modified for different spectrometers and light sources, and
- The correction is the best possible and does not compromise the spectral performance.

Disadvantages

- The attachment will eat ~40% of the entering light,
- It increases the footprint of the system,
- It is an additional optical instrument, and
- It adds to the cost.

4.1.6.2 Internal imaging correction by toroidal mirrors

Internal correction is offered by three different methods, the most popular of which uses the collimating and/or focusing mirror for beam deformation. If the mirror is not spherical but toroidal, it will correct the different pathlengths of tangential- and sagittal-oriented rays. If the correction is performed at only one mirror, the difference in curvature is stronger than if both mirrors are toroidal. Toroids provide a ring-shaped, different curvature over the surface of the mirror that is superimposed on the basic spherical curvature. Unfortunately, one curvature disturbs the other, inhibiting a perfect solution. How well the correction works depends on the focal length and the angles of the spectrometer.

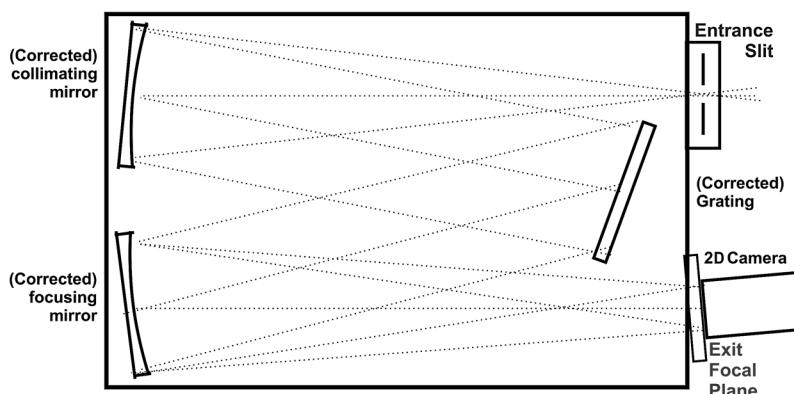


Figure 4.10 Imaging spectrometer with internal optical correction.

Advantages

- The number of components, and therefore the losses by reflection, remains unchanged; and
- The outer shape, the entrance, and the exit of the instrument are similarly unaffected.

Disadvantages

- The correction will not be perfect but rather a compromise: neither the spectral nor the imaging performance will be optimal; and
- Toroidal mirrors are rather expensive (replacing them with standard spherical mirrors is possible).

4.1.6.3 Internal imaging correction by a curved grating

The grating can be curved vertically to correct the pathlength of x and y beams while leaving the mirrors spherical. This method is only possible for a specific working angle of the grating, which must be tailored to the spectrometer; however, for that angular position, perfect correction can be attained. Note that turning the grating will change the optically effective angle of curvature. For spectrographs with a fixed grating position, this is an elegant and useful solution, but not for variable spectrometers. Curved gratings need extra design, making them expensive for small volumes. To remodel an existing system from a curved grating to a straight grating is very easy. For obvious reasons, the method was not established.

4.1.6.4 Internal imaging correction by a Schmidt corrector

A Schmidt corrector (or Schmidt plate) can be reflecting or transmitting. It was invented to correct image fields for large astronomical cameras in the 1930s. The surface is slightly cone shaped, which corrects the ray travel in both orientations; the result is a “flat field” distribution that also is desirable in photography and electronic imaging. As in the external case, the Schmidt corrector separates the functions of the mirror and the corrector. Figure 4.11 illustrates a reflective plate mounted halfway between the grating and focusing mirror. The corrector only works with parallel and reproducible illumination. Besides perfect illumination of the entrance, this design also requires the grating to turn on-axis, as marked by the dot. The camera is moved to the “rear end” of the spectrometer. It would not be a problem to add another entrance and slit exit, or a camera exit.

Advantages

- Very good correction for both spectral and imaging applications,
- Only one additional optical component,
- A footprint similar to noncorrected systems, and
- The Schmidt plate can be replaced with a plane mirror if pure monochromator functionality is needed.

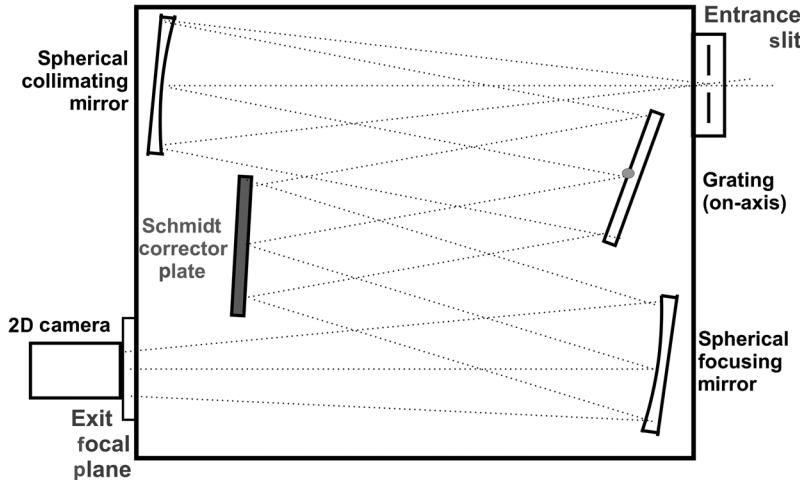


Figure 4.11 Imaging spectrometer with a Schmidt corrector.

Disadvantages

- The technique requires on-axis rotation of the grating and perfect entrance illumination,
- The corrector is a rather expensive device, and
- It only works in the pre-defined setup.

4.1.7 Prism spectrometer

A grating spectrometer can be converted to a prism version if the disperser rotates around the pivoting point of the dispersing element. However, the progress of the turn is not a sine or cosine. The output wavelength and dispersion of a prism spectrometer depend mainly on three parameters: the refractive index (RI) at the given wavelength λ , the prism's internal angle A , and the illumination angle i_1 . The wavelength or wavenumber function will follow a polynomial curve. The resulting dispersion depends on the focal length.

In a single disperser instrument (Fig. 4.12), a prism or grating can be used and turned around the pivoting point (dot). In the figure, a “Littrow prism” is used; this version has a cut end at working angle A . Shortening the prism end has no disadvantage because it is not used, but the “nose” can interfere with the beam. The cut also offers the advantage of increasing the base length b , which is important for the resolution of the prism, without increasing the other sides. A spectrometer with certain internal angles and a pre-produced driving system (disk or control program) is only useful for one type of prism. The only variable left, then, is the focal length. This complication is a large reason why prism spectrometers went out of fashion. A grism can be used instead of a prism, requiring a different, more-complicated turning algorithm.

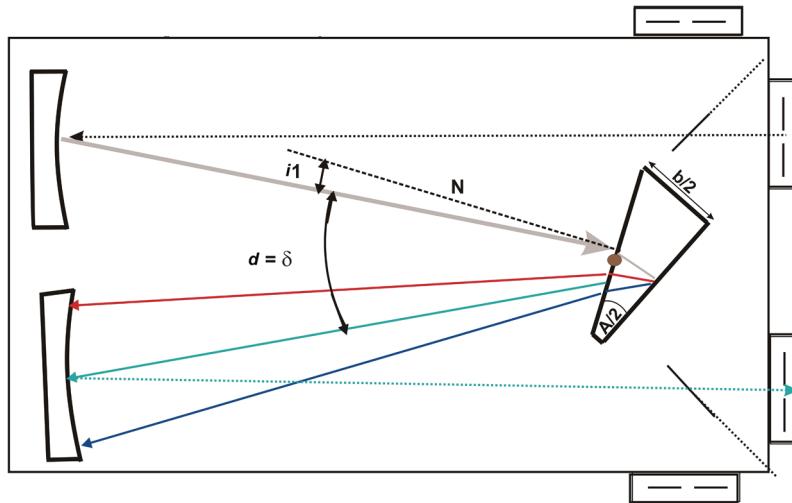


Figure 4.12 Beam travel in a typical reflecting prism spectrometer.

4.1.8 Dispersion of a prism spectrometer

Equation (3.2) leads to d_{\min} , the minimum deflection “ d .” The dispersion is found by the difference of d_{\min} at two neighbor wavelengths. Because the prism will be turned, Eq. (3.3) is used (see Section 3.4). The dispersion RD at the exit results from the differential application of the equation. The dispersion of a prism spectrometer and that of a grating based system is calculated the same way [see Eq. (3.4)], which allows, in conjunction with the focal length, the stepwise calculation of dispersion. At a certain position in the output, the valid equation is the reciprocal dispersion, shown in Eq. (3.5).

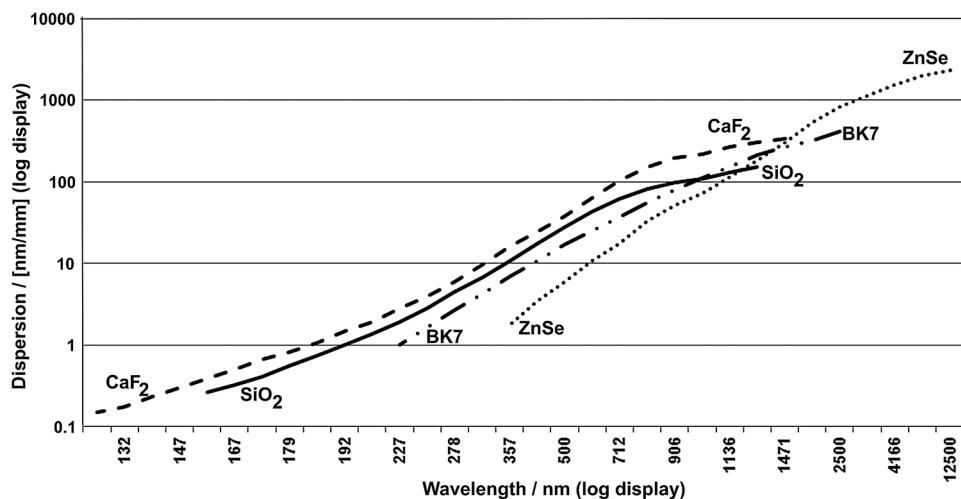
In a scanning monochromator or a flexible spectrograph, the sum of the angles $\varepsilon_1 + \varepsilon_2$ composes the inclusion angle δ , which must be equal to d of the prism, to guide the required refracted light to the exit. Table 4.1 presents the RI of some popular materials, and Fig. 4.13 depicts the data as typical dispersion curves.

The curves are based on prisms, which are either reflecting at 30 deg or transmitting at 60 deg, mounted in a 1-m spectrometer at $\delta = 20.5$ deg. For a better overview, both dimensions (wavelength and dispersion) are shown in a logarithmic manner.

The illumination angle i_1 (or α) plays a very important role, as it does in grating applications, too. The internal angles ε_1 and ε_2 of the spectrometer have a stronger effect compared to a grating system. Almost all materials useful for prisms show an increase of the refractive index n with increasing photon energy (eV) with respect to declining wavelength (nm). Besides the RI behavior, the absorption constant k limits the useful range. As soon as k departs from the zero value, the prism is not useful anymore because interference and losses will occur. For many materials the phenomenon of $k > 0$ appears in the IR and UV directions, as well. The useful range of a single prism is much wider than that

Table 4.1 Refractive index of materials.

Photon energy [eV]	Wave number [cm^{-1}]	Wavelength [nm]	RI SiO_2	RI BK7	RI CaF_2	RI ZnSe
0.10	800	12,500.000	—	—	—	2.4100
0.20	1,600	6,250.000	—	—	—	2.4230
0.30	2,400	4,166.667	—	1.4800	—	2.4330
0.40	3,200	3,125.000	—	1.4850	—	2.4370
0.50	4,000	2,500.000	1.4300	1.4900	—	2.4400
0.75	6,000	1,666.667	1.4427	1.4950	1.4260	2.4450
1.00	8,000	1,250.000	1.4476	1.5000	1.4275	2.4650
1.25	10,000	1,000.000	1.4505	1.5050	1.4290	2.4770
1.50	12,000	833.333	1.4539	1.5100	1.4306	2.5100
2.00	16,000	625.000	1.4584	1.5155	1.4331	2.6100
2.50	20,000	500.000	1.4626	1.5217	1.4360	2.7000
3.00	24,000	416.667	1.4687	1.5292	1.4410	2.8116
3.50	28,000	357.143	1.4762	1.5383	1.4460	3.0600
4.00	32,000	312.500	1.4832	1.5495	1.4522	—
4.50	36,000	277.778	1.4958	1.5634	1.4590	—
5.00	40,000	250.000	1.5086	1.5812	1.4680	—
5.50	44,000	227.273	1.5238	1.6050	1.4780	—
6.00	48,000	208.333	1.5430	1.6381	1.4900	—
6.50	52,000	192.308	1.5670	1.6854	1.5040	—
7.00	56,000	178.571	1.6000	—	1.5227	—
7.50	60,000	166.667	1.6430	—	1.5670	—
8.00	64,000	156.250	1.7020	—	1.6280	—
8.50	68,000	147.059	1.7830	—	1.7080	—
9.00	72,000	138.889	—	—	1.7780	—
9.50	76,000	131.579	—	—	1.8500	—
10.00	80,000	125.000	—	—	1.9140	—

**Figure 4.13** The principal dispersion behavior of prism spectrometers using the materials listed in Table 4.1.

of a single grating, with the exception of Echelle gratings. At given n values, the dispersion improves with the internal prism angle A . It follows that the dispersion can be changed by changing the angle A so long as the value is between 10 deg and 90 deg. Unlike the grating, a prism always works in transmission mode (otherwise no refraction would occur). For reflective spectrometers, the prism's rear side is coated by a reflecting layer. A reflecting prism of $A/2 = 30$ deg has the same effect as a symmetric transmitting prism of $A = 60$ deg. If a grating is attached to the prism, the combined function becomes that of a grism.

Prisms were very popular until the 1970s, mainly in spectrometers for ranges above 10 μm or below 200 nm. They were used in the IR because of the wider working range and the high refractive indices of available materials, such ZnSe, that produce good resolution. They were used in the vacuum UV because of the remarkably better efficiency and better dispersion compared to gratings. With the arrival of holographic gratings providing higher line densities ($\geq 3600 \text{ mm}^{-1}$), and better coating techniques, gratings have closed the gap. As a consequence of the increasing use of area detectors, prisms lost value because the nonlinear dispersion over wider wavelength ranges is often a disadvantage. One significant characteristic of prisms remains: they always provide unquestionable wavelength information; no order sorting or extra filtering is required. That trait, in turn, makes prisms a well-suited variable filter in front of a grating spectrometer (as discussed further in the next section).

Figure 4.14 shows a typical dispersion function of a prism spectrometer. The vertical axis represents the reciprocal dispersion in a logarithmic manner. The gradient of the curve changes twice: In the deep UV, there is a dramatic change due to the strong change of the refractive index n . In the near-IR the gradient changes again because n almost reaches a plateau. Between 1000 nm and ~ 2800 nm, the RD will almost stay flat. Note that this is a typical curve; variations are due to specific n/k data.

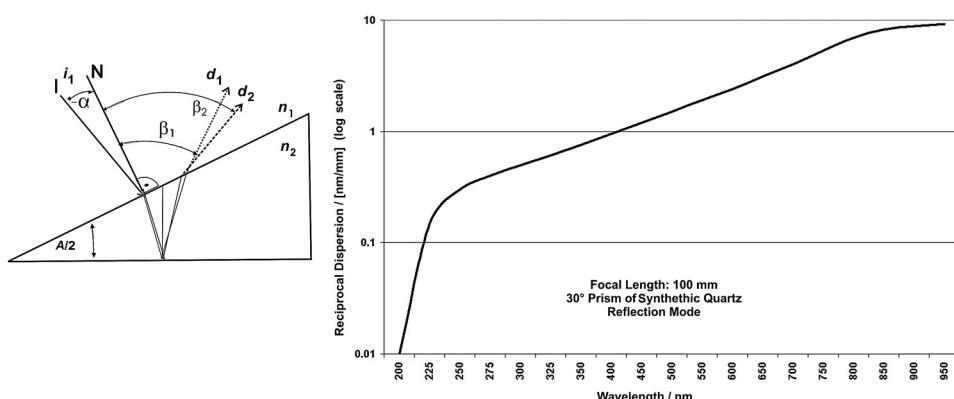


Figure 4.14 Reciprocal dispersion of a quartz prism with a 30-deg angle, working in a spectrometer with a 100-mm focal length.

4.1.9 Echelle grating spectrometers

In general, prism spectrometers are very interesting in combination with Echelle gratings. A typical example for systems with dual dispersers are Echelle-based spectrometers, which use the higher orders to increase resolution and benefit from the wide working range of Echelles. Echelle monochromators and spectrographs show distinct requirements in construction and application. They are treated in detail in Section 4.4.

4.1.10 Transmission spectrometers

Transmission spectrometers are described merely for the sake of completeness because they do not fulfill the requirement of flexibility. The principal design has clear advantages if the target wavelength interval is short. The typical range of a single setup is approximately three spectral orders. In practice, the application parameters are always fixed, and the aperture plays an important role. The ideal transmission spectrometer would have an “inline configuration,” where the entrance transfers the light to a large collimator. For short intervals of λ , this can mean a corrected lens, but a commercially available lens system with high throughput is generally used; herein lies the first limitation. It is difficult to find a commercial lens system reaching below 300 nm and above 2000 nm. The aperture, on the other hand, can be as large as $f/2$ at focal lengths up to 200 mm. Collimating and focusing optics are almost always the same type. If the dispersion element is a prism, the output will have a nonlinear dispersion axis. Again, for comparison, a transmission grating needs to fulfill several requirements to work properly: the substrate should have an absorption factor of $k = 0$ in the intended range to avoid interference. Also, no interference should occur between the substrate and the structure; this can be realized if the structure is ruled or burned into the substrate directly, which is possible with several glass or plastic materials. Alternatively, a very thin foil can be bonded to the substrate if there is a close match of n and if $k = 0$ for the substrate, foil, and adhesive. These parameters will again limit the range to about 300–2000 nm. It makes no difference as to which side of the grating the structured surface is applied—a refractive component will be added to the grating function anyway because there is no way to guide the light orthographically through the substrate and get useful diffraction.

After selecting all of the components, the internal angles of the spectrometer will result from the calculated working angles. The resulting instrument will only allow small variations. At the exit, it will be useful to track only one wavelength, mount several discrete detector elements for different wavelength ranges, or mount a chain or area detector to monitor a certain interval. It is a logical consequence that transmission spectrometers are mostly designed for closely defined, industrial applications and rarely for research labs with changing requirements. The equations for the grating and prism solutions remain known. The total length of the light path is calculated

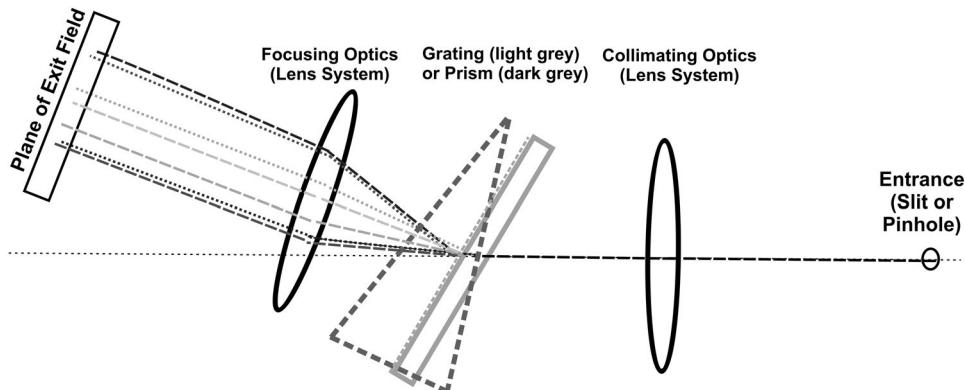


Figure 4.15 A simplified but realistic transmission spectrograph.

a little differently compared to a reflecting spectrometer. The dispersed (exit) arm begins at the rear side of grating or prism; Fig. 4.15 displays this arrangement, but only for the grating, not for the prism.

In reality, the single lenses shown are whole lens systems. It does not matter whether a grating or prism is used—the illuminating beam must hit the disperser under a certain angle other than perpendicular in the horizontal plane. The dispersion defines the mean output angle, which is the angle of the output leg relative to the horizontal plane. In most cases, an area detector is used. It is shown in Fig. 4.15 that (a) the orientation of the output wavelengths are inverted between grating and prism, and (b) the grating dispersion is close to linear, but that of a prism is nonlinear. The dotted lines show the beam travel of a grating, and the prism beam is the dashed line. It is obvious that the output arm of the prism-dispersed light starts more left than shown in the figure because of the different thickness of the elements. The fine tuning of detector and angle of the leg is possible within certain limits. The transmission spectrometer, as shown, can be considered single axial because the angles result from refraction or diffraction, and not from optical deviation. If a grism were designed for a single angular position, it might be possible to find a solution for a single straight axis, as demonstrated in Section 3.5.

4.2 Grating Rotation and Actuation

From the grating equations, it is evident that the positioning follows right-angled triangular functions. If the photon energy is entered as wavelength λ , the coherence is a sine function calculated by the relation $\sin = (\text{length of opposite side})/(\text{length of hypotenuse})$. Because the zeroth order lights up at $\sin = 0$, it is an easy and useful calibration point (see Chapter 7). If, on the other hand, the energy is presented as frequency ν or as wavenumber $\tilde{\nu}$, the coherence becomes a cosine function: $\cos = (\text{length of adjacent side})/(\text{length of hypotenuse})$. The reference point would then be $\cos = 1$, and the numerical

value would be infinite. This behavior makes it difficult to calibrate, and positioning in the vicinity of the reference position would end in large numerical changes per degree, which is why the sine drive has been used since the early 1930s. It also is the major reason why grating spectrometers use the wavelength as the parameter of photon energy instead of the more sound wavenumber, GHz, or eV scale (those are driven by the cosine function, the “cosecant”). The cosecant drive is very useful in angular ranges other than the reference point, such as the IR range. If, for instance, a spectrometer with a 150-mm^{-1} grating works between $2\text{--}10\ \mu\text{m}$, or $1000\text{--}5000\ \text{cm}^{-1}$, the grating runs between 8.5 deg and 47 deg. The cosecant will realize a $\tilde{\nu}$ -linear scan.

There is one application in the UV–Vis range that calls for the cosecant: Raman spectroscopy. Because Raman signals appear in a certain, reproducible distance of energy from the excitation source, the $\tilde{\nu}$ -linear scan is the method of choice. Pure electromechanically driven systems face the problem that the cosecant must be aligned and calibrated to the photon energy of the excitation laser, which leads to the situation that the excitation is documented in nm, whereas the Raman spectrum is recorded in cm^{-1} . Although sine and cosine drives were state of the art up to the early 1990s, circular driving systems can emulate both functions via stepper motor systems with software control. When it comes to classical actuation, this book concentrates on the sine drive.

4.2.1 Classical driving system

In the classical grating drive system (Fig. 4.16), a worm screw pushes a lever that turns the grating. The micrometric worm screw is the short side of a triangle, and the lever (the “sine arm”) is the hypotenuse. The linear movement, actuated by

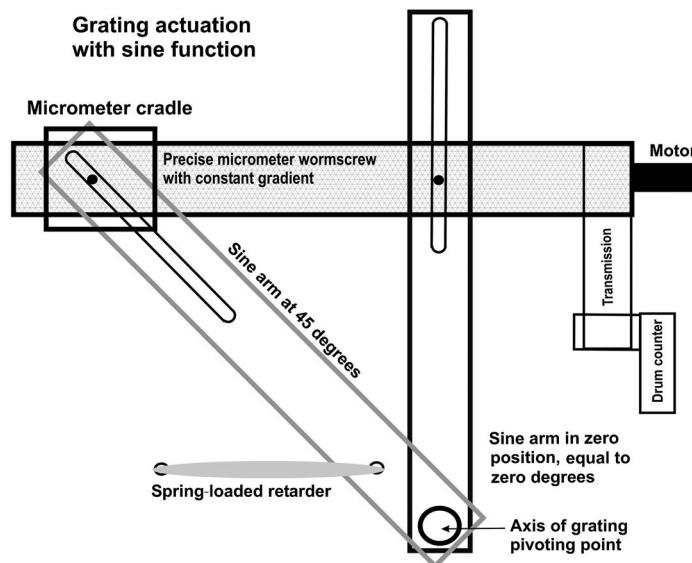


Figure 4.16 The classical “sine drive.”

the worm screw, is turned into a sine function, placing the pivoting point in the grating's active surface. The worm screw moves a sliding carriage that is mounted around the screw by an almost-backlash-free joint. The carriage itself pushes the sine arm through a precise ball/polished face sliding surface. The sine arm is held by a pre-loaded spring. The goal of the construction must be to realize long-lived systems that are reliable and free of squeezing and friction. The worm screw, in most cases, is moved by a fine stepper motor. A transmission moves a drum counter to show the position. The system consists of materials with minimal abrasion and a low thermal coefficient (i.e., bronze or Super Invar[®]), and effect very little angular change per turn: typically 0.1 deg/turn. In Fig. 4.16, the sine arm is shown for two positions: 0 deg in the center position, and 45 deg in the diagonal. It can move in both directions. The limit switches and position encoder are not shown. If the system should present wavenumbers, the function would be a cosine instead of the sine.

Switching between different energy axes (such as λ and ν) was not possible before the availability of microprocessors. A spectrometer was only calibrated for one type of energy and optimized for only one grating. The worm screw had to be extremely accurate to provide linearity. The screw was moved by hand or by DC motor (and by stepper motors since the 1970s). Calibration was done at two points and was purely mechanical. It is easy to imagine the efforts to adjust and synchronize a double or triple spectrometer. Today, we have access to highly resolving stepper-motor systems that provide more or less perfect reproducibility that can correct hysteresis (backlash) effects between forward and backward scanning. High-performance processor systems save the specific calibration data for several gratings, entered by the manufacturer or the user, in nonvolatile memory.

Since about 1990, spectrometers with direct drive of the grating table are available. The switch between different energy axes in some systems is very easy and safe. The synchronization of several spectrometers or stages is often provided. Even the field upgrade of gratings or combination of spectrometers is seamless if the right manufacturer was selected at purchase. Spectrometers with more than one grating on a table use stop-less turning stepper-motor systems. The electronics and software then have to convert a rotation into a sine (or cosine) function for two or three gratings, which is not trivial.

4.2.2 Grating actuation by a rotating system

In order to increase flexibility, spectrometers utilizing multiple gratings have been on the market since the end of the 1980s. Most manufacturers utilize three gratings on a rotating table. Different solutions have been presented; Figure 4.17 shows two versions.

Figure 4.17(a) depicts the off-axis version. The table rotates around its own center, and the surface of the active grating travels sideways through the beam while turning. Figure 4.17(b) depicts the on-axis version. The table is mounted so that the active grating rotates around its surface center. The table

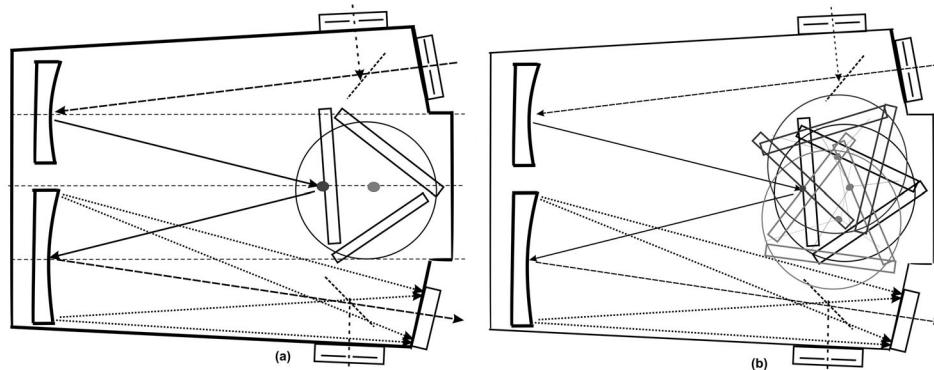


Figure 4.17 (a) Off-axis and (b) on-axis version of a triple-grating spectrometer.

then moves sideways and partially blocks the beam if the width of the system is not increased.

Off-axis version

A multiple-grating configuration, turning around the center of the table, is marked by the light-grey dot in the left side of Fig. 4.17. No matter which direction the device turns, the center of the grating (dark-grey dot) will move out of the center of the beam. To compensate for the loss in surface, wider gratings can be selected to catch the illumination light; however, this can disturb the beam, especially in the area of the field output. The construction needs to account for that issue. Eventually, the system must become wider than initially assumed.

The pure spectrometric features (dispersion, resolution, calibration, etc.) will not principally suffer from the off-axis move, whereas imaging applications can because the sagittal rays can change the pathlength as the grating rotates.

On-axis version

The right side of Fig. 4.17 shows a triple-grating configuration that pivots around the center of the active grating surface (dark-grey dot). To realize this arrangement, the surface and turning behavior are identical to that of a single-grating solution, but the mechanical construction is far more complicated and susceptible to failure, compared to the off-axis version. The whole mechanical system must rotate around the dark-grey dot. Rotating the grating towards the entrance will swing the table towards the exit, whereas turning the grating towards the exit will swing the table towards the entrance. The black drawn grating arrangement shows the grating near zero position, the dark-grey arrangement at about 45 deg towards the entrance, and the light-grey position shows the situation under a 45-deg turn towards the exit. The expense can either be limitations in the output field or even more extended angles.

Dual- and quadruple-grating configurations are also available. It is easy to imagine that a twin system with gratings back-to-back does not differ much from a single-grating system—the off-axis shift will be small—whereas a four-grating system can require even wider angles than a triple-grating system.

Clearly, the optimal situation involves the grating turning in the center of its reflecting surface. Single-grating instruments normally realize that scenario. For tables with several gratings, four basic constructions are available. Note that construction 1 relates to an off-axis version, and constructions 2–4 relate to on-axis versions:

1. The table is in a fixed position and rotates around its own center. Thus, the gratings do not rotate on their active surface; rather, they move through the beam (how much depends on the distance from the surface-pivoting point).
2. During the change of gratings, the table is moved out of its holder and then returned after indexing the gratings. The grating rotates on the active surface.
3. A dual-axis stepper motor rotates and moves the table back and forth while the grating turns. This behavior will keep the grating in the optimal position.
4. Instruments holding four gratings have been designed with index holders like dice. The gratings are mounted under 90 deg, each on the circumference in a vertical manner. The active grating is moved to the front position by one motor; a second motor turns the system around the pivot of the active grating. Besides the complexity, this solution does not increase the internal angles of the spectrometer, but because the holder rotates in the vertical axis, the system needs to be higher than the optics alone would require.

As gratings receive and send collimated light, there is no imperative necessity to rotate in the surface. All rules and functions apply, even if the grating moves to one side with rotation or changes the distance to the mirror slightly. It will lead to small deviations of α , β , and ϵ . Handling that behavior is only a matter of the processor system and its firmware. A sideways movement introduces a declining projected grating area and requires compensation. Consider an example with a 70-mm-wide grating: Turning the grating moves it to one side. The width is simultaneously reduced due to pure rotation. If the grating's surface is 30 mm off axis, the grating will move 26 mm sideways while the working angle changes from zero deg to 60 deg. The total loss is up to 50% of the area; however, if the grating turns towards the entrance and ϵ_1 is 15 deg, it will have full projection at 15 deg, and the maximum loss will be only 30%. If the installed gratings are 84-mm wide, all losses are compensated, at which point the gratings will always provide full aperture and work fine.

If the construction is based on versions 2–4, the grating must move in at least two axes. These solutions require a rather complicated mechanism with enough longevity to provide the reliability and steadiness needed for the lifetime of the instrument.

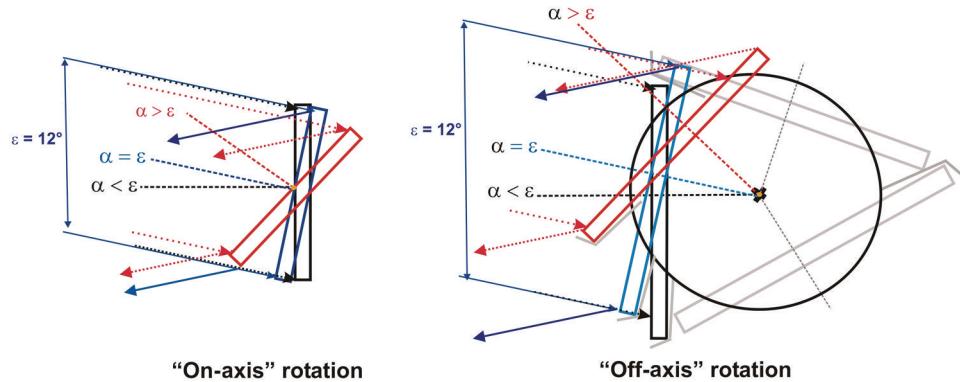


Figure 4.18 Influence of the grating rotation on the active surface.

In Fig. 4.18, two spectrometers with identical internal angles ($\delta = 30$ deg and $\epsilon = 15$ deg) demonstrate the impact of rotation on the grating aperture. On the left is a grating pivoting in the front surface. It demonstrates that the illuminated surface, and hence the aperture, drops with increasing working angles. An on-axis system would behave identically. On the right is a system with three gratings, and the pivot point is the table center. The gratings are peripheral, and they rotate and move to the side. The “grey gratings” are only shown under zero deg; the baffles avoid reflection from the side of the substrates. Turning the single (left part) grating to 45 deg results in a surface loss of 30%, whereas the triple system faces a loss of 50%. In other words, for a constant aperture, the gratings in the triple system need to be $\sim 30\%$ wider.

Grating actuation, steering, and calibration are discussed further in Chapter 7.

4.3 Multiple-Stage Spectrometers

The requirement for high resolution (after the variations of gratings are resolved) can only be fulfilled by increasing the focal length. In the case of atomic spectroscopy, where sub-picometer resolution is required, this leads to focal lengths of ≥ 2 m. If the system will be alternatively used as a monochromator and a spectrograph, the monochromator resolution can be suitable for a particular experiment, but the dispersion in combination with the size of the area detector produces a spectral interval that is too short. The requirement can be solved by using multiple gratings with different dispersion or the translation of the output image. The image plane of the area detector can be reproduced outside the device, coupled with transfer optics, which allows switching between 1:1 and, e.g., 1:5 reduction. A larger interval is available but at the expense of resolution. Even at 1:1, aberration will be increased by the transfer optics, stray light will increase, and transmission will drop. As described in Section 2.6.5.4, the majority of area detectors deliver resolution performance below that of the monochromator version. Because optical transfer systems can

magnify by, e.g., 3:1, as well, the same resolution can be reached with both modes (monochromator and spectrograph). Folding the beam several times can extend the focal length without increasing the footprint. Naturally, this will increase reflection losses and stray light.

4.3.1 Double-pass spectrometers

A probable method of increasing focal length sends the beam through the spectrometer twice by adding a double-pass option. The dispersed beam does not leave the system, but rather it is sent back to the entrance and travels the whole path once more. Figure 4.19 displays the double-pass arrangement. To work properly, it must fulfill certain requirements. The height of the entrance and exit must provide enough room to completely separate two beams. This, in turn, calls for stigmatic transfer, either by curved slits or internal image correction. In practice, the useful height of the slits should not be less than 20 mm. One half of each slit (the upper part of the figure) will be blocked by a diverter mirror. The lower part will be the entrance and exit, respectively. The beam will then travel slightly inclined and hit the center of the grating. After dispersion, the beam will still travel slightly upward—it will leave the output in the upper section. However, before it reaches the exit, there is a diverting mirror that will send the light to its counterpart behind the entrance slit. The two diverter mirrors are placed at a distance from the spherical mirror, which creates a virtual slit plane between them. This arrangement is usually between the grating and the front wall, behind the center of the grating. Because the beam arrives above the center, the second diverter sends the already-dispersed

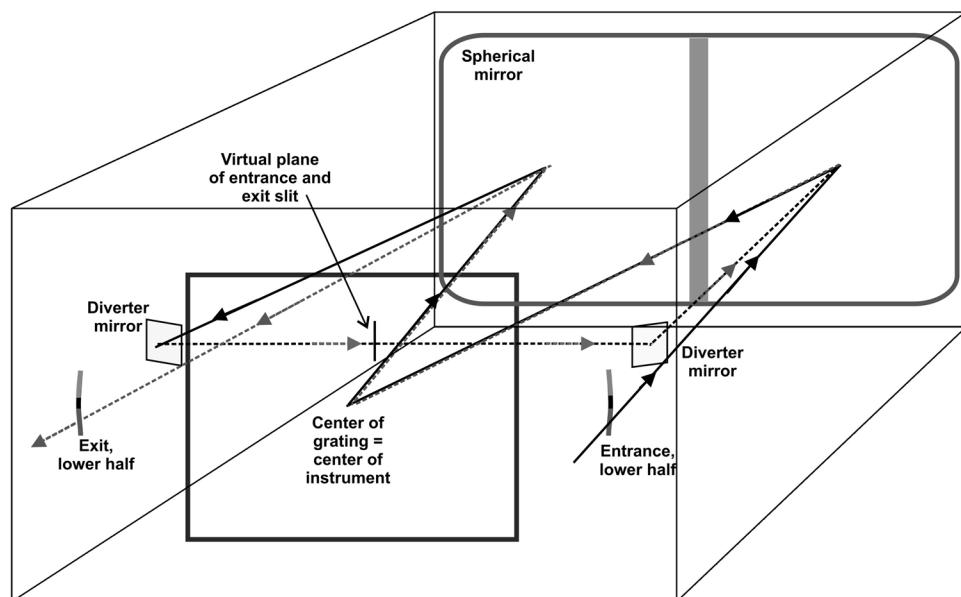


Figure 4.19 Beam configuration of a double-pass spectrometer.

light on another tour through the system. The light then travels slightly tilted downward, hitting the center of the grating again.

Finally, the double-dispersed light leaves the exit (slit or area) at the same height that the entrance received the incoming signal (in this case, the lower half). The first tour of the beam is marked solid, and the second is dashed. The used parts of the slits are dark grey, and the blocked parts are marked black. To realize its proper function, the spectrometer must provide full symmetry and stigmatism. Whether the beam passes in front or in back of the grating depends on the ratio of the length and width of the spectrometer. Halfway between the two folding mirrors is the virtual focal plane. A variable aperture can be placed there to help reduce stray light if the beam passes behind the grating. A system prepared for double-pass operation should provide an easy and reliable activation of the double pass by moving it up and down at a single carrier, requiring adjustment only once. Regardless, the entrance beam and detector position will require readjustment whenever the double pass is moved in or out. The double-pass option is an effective tool to double dispersion and resolution. The light flux will suffer because only half of the slit height or even less is available. Because the beam passes twice, the transmission will be halved. The stray light will be improved remarkably.

The dispersion of double-pass spectrometers (RD-DP) doubles compared to the single-pass version:

$$RD-DP = \lambda/[2(2f \times \tan \phi)], \quad (4.1)$$

and the result is half of the numerical value.

4.3.2 Double spectrometers

The best way to radically improve the contrast is to configure two or even three spectrometer stages in series. To make that work, two stages are coupled together, and the exit (slit) of stage one becomes the entrance (slit) of stage two. The second stage is only fed by the spectral data pre-selected in stage one, and will further disperse and clean the beam. In short, this arrangement produces an exponential yield in contrast C plus twice the dispersion and increased resolution. The contrast ratio of a double spectrometer can hardly be matched by any single-stage system; it is far more than adding the two C values. Consider the following example as proof: A certain wavelength and bandwidth is required at a contrast of 10^4 . An available single-stage spectrometer provides the wavelength and bandwidth at a stray light level of 1%, producing a C of 10^2 . If an identical spectrometer is used in double configuration, the contrast will be at least 10^6 . In terms of dispersion, there is no need to use double systems so long as the focal length of each stage does not exceed ~ 0.75 m. The resolution will not completely double, anyway.

To reach optimal resolution, perfect synchronization of two instruments, two gratings, and three slits is necessary. This scheme requires considerable effort and a highly stable environment in terms of vibration and temperature.

Multi-stage spectrometers are very sensitive to any kind of asynchronism because it can shift apart the center wavelengths of the stages. If that happens, the transferred spectral band will become asymmetric or even be cut on one side, the signal transfer will suffer, and the displayed spectral position will change. It is typically required that the center wavelengths of the stages stay within 1/22 (see Section 2.6.5.1) of the bandwidth transferred. To achieve that parameter, the wavelength drive—whether mechanical by spindle or by discrete stepper motors—should resolve better than 1/200 of the bandwidth. If the system is exposed to changes in temperature, the thermal coupling needs to ensure that no internal differences and no torque occur.

In summary, it is not enough to put single spectrometers together. Double spectrometers are mainly used for contrast enhancement, as in radiometric applications, or in the production of spectrally pure light over wide wavelength ranges. Raman and fluorescence are two typical analytical methods for multiple-stage systems. Two inversely dispersing spectrometers compose a subtractive system, allowing special applications.

4.3.2.1 Subtractive spectrometers

Subtractive spectrometers have the ability to provide a defined bandwidth homogeneously over an area. The distribution can be of identical power and wavelength throughout, which offers two attractive applications: (a) homogeneous illumination of an area, and (b) a third stage can assume the pre-selected and mainly distortion-free (stray-light-free) spectral interval, delivered in one spot or a narrow slit. The third advantage applies to time-critical experiments in the sub-nanosecond range because subtractive spectrometers provide the same traveling time for all wavelengths reaching the exit. The spectrometer can then supply an area detector with well resolved and clean light signals. By definition, double and triple spectrometers are not required to have the same focal length or grating. Indeed, deviations in entrance and exit angles, and the resulting differences in slit widths, need to be taken into account. Otherwise, exorbitant losses of light can result if incorrect settings are applied. Multi-stage systems will always be defined after the application requirements, making it provide more specific advice.

The dispersion of double spectrometers is found to be additive and subtractive, respectively:

$$RD-DA = \lambda / [(2f \times \tan \phi) + (2f \times \tan \phi)], \quad (4.2)$$

$$RD-DS = \lambda / [(2f \times \tan \phi) - (2f \times \tan \phi)]. \quad (4.3)$$

In practice, the numerical value of an additive, symmetric setup will be half of the value of the single-stage system. A perfect subtractive system will produce a zero in the denominator; in the special case, that should be interpreted as “no dispersion.”

Figure 4.20 displays a classic double spectrometer used in additive and subtractive setups. In the former mode, the second stage is entered through its side-entrance slit; in the latter mode, the beam leaves the first stage and re-enters

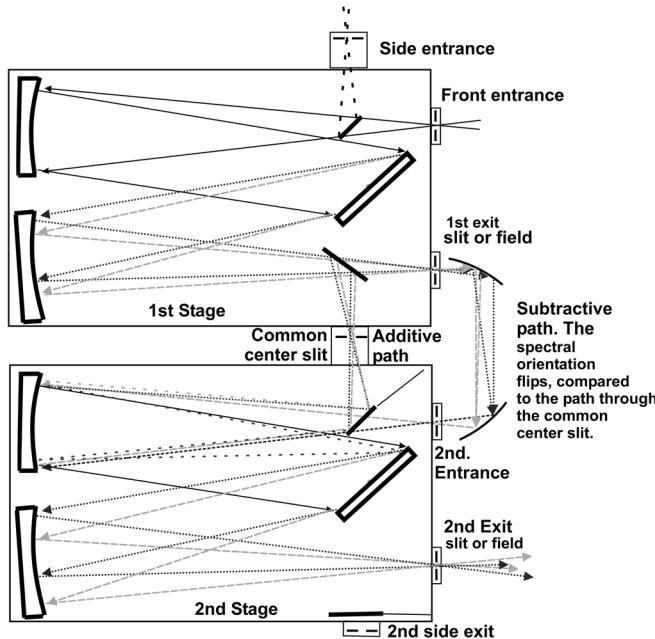


Figure 4.20 A classical double-spectrometer setup, providing additive or subtractive operation.

Table 4.2 Comparison of different focal lengths for double-pass and double spectrometers.

Parameter	1-m Single	2-m Single	1-m Double Pass	1-m Double Stage	Scale
Dispersion	0.788	0.396	0.384	0.384	nm/mm
Resolution	50,000	110,000	95,000	85,000	R_r [numerical ratio]
Slit width	1.27	2.525	2.54	2.54	mm
Slit height	20	20	8	20	mm
Luminosity	49	48.4	15.7	98	L [numerical]
Contrast	10^5	10^6	3×10^5	10^{12}	C [numerical ratio]

the second stage after an external mirror system flips the beam orientation. In the figure, the black dotted rays indicate a shorter wavelength (i.e., blue), and the grey dashed lines indicate a longer wavelength (i.e., green). The rays are shown for both modes and portray the beam inversion by the external path. The details of additive and subtractive operation are described in Section 4.3.3.

Table 4.2 lists typical values for a grating of 1200 mm^{-1} with the following parameters: $110 \text{ mm} \times 110 \text{ mm}$ illuminated area, tungsten halogen source, 0.1-nm bandwidth at 500 nm, monochromator mode, and resolution at minimum slit width. Except for the limiting resolution, the double monochromator is rated the best due to the wide slit width for the required bandwidth (with the best contrast by far). Because the slit area contribution is squared to the luminosity L , it dominates the transmission, which becomes half. The comparison can be repeated for any focal length; the result will tend to be the same.

4.3.2.2 Efficiency behavior and analysis

The wavelength-dependent efficiency T is mainly influenced by the grating. For the mirrors, the general assumption for wavelengths above 190 nm is a reflection efficiency of 0.88–0.96 per mirror. The exact value depends on the wavelength and coating. If these are not known, 0.9 is a good guess. For gratings, it can be estimated that at the optimal wavelength the dispersion efficiency can reach up to 0.9. The reflection losses of the front face depend on the coating and will be in the range of 3–15%, resulting in a total efficiency of about 0.85. Taking into account the efficiency of 2–4 mirrors, the total transfer efficiency T of spectrometers in the UV–Vis–NIR range will be 0.4–0.7. It is no mistake to estimate a total T of 0.5 at blaze wavelength. Apart from the blaze wavelength λ_1 , the grating will drop towards the UV, reaching 50% of the maximum at approximately $\lambda = 0.7 \times \lambda_1$. The same value will probably be found at $\lambda = 2.0 \times \lambda_1$ towards the IR. In both cases, the total T is ~ 0.25 . The interval between them is generally seen to be the optimal working range within one order (this does not consider changes in polarization and anomalies). All of the effects are found twice in a double-pass or double spectrometer; thus, the optimal range will be shortened. Measuring and analyzing the real T at a wavelength or for a certain range can be difficult—see Section 7.3 for further advice.

4.3.2.3 Energy transmission and bandwidth of single-, double-, and triple-stage spectrometers

As shown in Fig. 4.21, the lateral energy distribution in the output slit of a spectrometer is a Gaussian curve. Adding stages modifies the behavior. Whereas the top of the curve becomes flatter, the edges become steeper. In fact, the bandwidth is reduced, and the resolution is improved. Superimposing

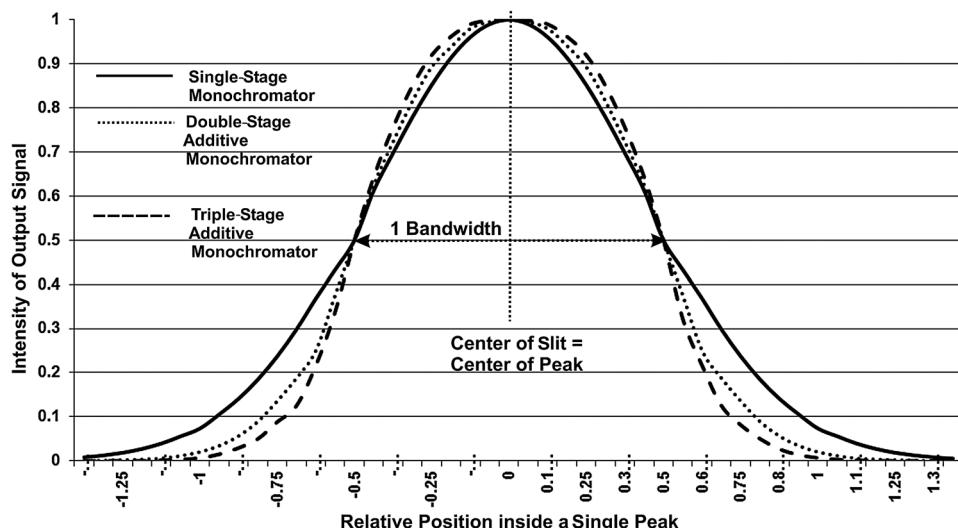


Figure 4.21 The energy transmission vs. slit position changes with the number of stages.

the normalized curves of single-, double-, and triple-stage monochromators illustrates the changes. Besides the vastly reduced stray light, the multiple-stage concepts also improve the steepness of the filter function.

4.3.2.4 Effects of photon traveling time (time of flight)

The traveling time that the light needs to move through an optical system, and the dispersion of the traveling time, will affect studies in the pico- and nanosecond time frame. In general calculations, only the time required to travel along the central paths of the x and y axes is calculated. The photons travel at (rounded) 30 cm/ns. If a setup comprises a light source, a 30-cm spectrometer (which alone has a path of about 110 cm), sample, and detection arm, a total light path of 2 m is easily reached, which produces a nominal traveling time of ~ 7 ns. The optical aperture adds extra time for those photons that travel along the outskirts of the multiple-cone-shaped lightpath (Fig. 4.22).

For example, a system can have an optical aperture of $f/4$ ($\omega = 0.005$, see Section 2.6.2.3). The cone spreads up to $1/8$ ($\sin 0.125$) of the focal length, realizing a side angle of 7.2 deg in all four directions. This arrangement produces an additional pathlength/time of flight of 0.8%. The distribution of the wings is not symmetric because only those photons that follow longer pathlengths contribute. The homogeneity is also influenced because most of the rays will stay near the center. If the grating stands under high angles, the rays hitting the nearer grating surface travel a shorter way than those in the center and at the far end. The “short-end photons” create an additional wing towards shorter times of flight, but because they affect the difference

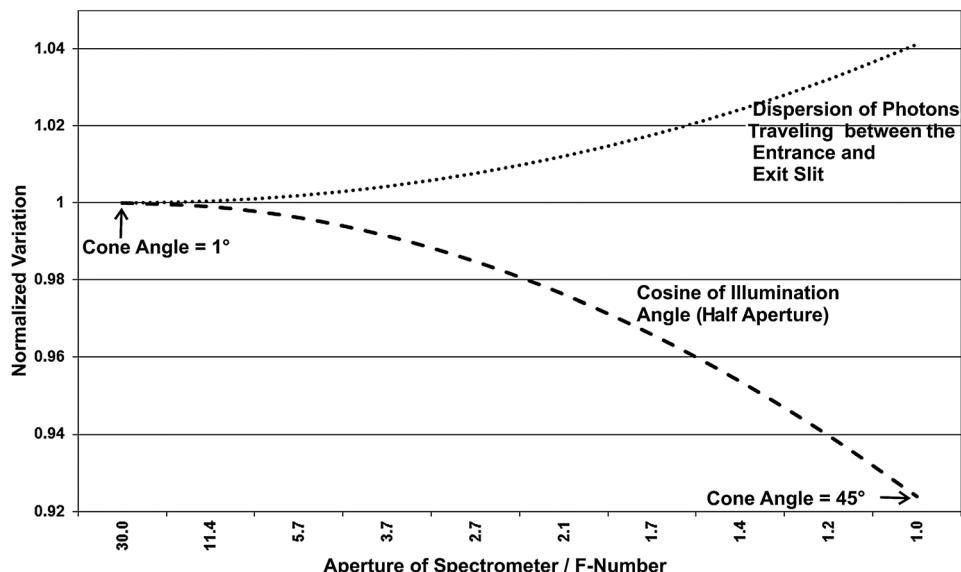


Figure 4.22 The dispersion of traveling time and the reduction of beam homogeneity versus the aperture.

in pathlength much less than the aperture does, the effect can be omitted. The same is true, in most cases, for time differences created by changes in the refractive index (i.e., lenses). Figure 4.22 shows that with increasing apertures, the time-of-flight distribution has a stronger effect and the inhomogeneity increases. For applications in the picosecond and nanosecond range, such as those analyzing pulsed laser signals or fluorescence lifetime, the impact might become critical. Aperture reduction is an appropriate tool to reduce the impact at the expense of throughput.

4.3.3 Construction considerations for double spectrometers

To investigate double spectrometers in general, a “ray tracing” is performed on the principal interrelations in beam guidance. Double spectrometers are often found as a tandem, with two parallel, identical instruments sharing one common, intermediate slit or aperture.

4.3.3.1 Additive setup

The simplified illustration in Fig. 4.23 shows an additive, parallel tandem. The beam guidance always displays one of the beam bundles, which is the most extreme one, but not the cones. To ease calculations, a grating of $k = 1 \mu\text{m}$ is

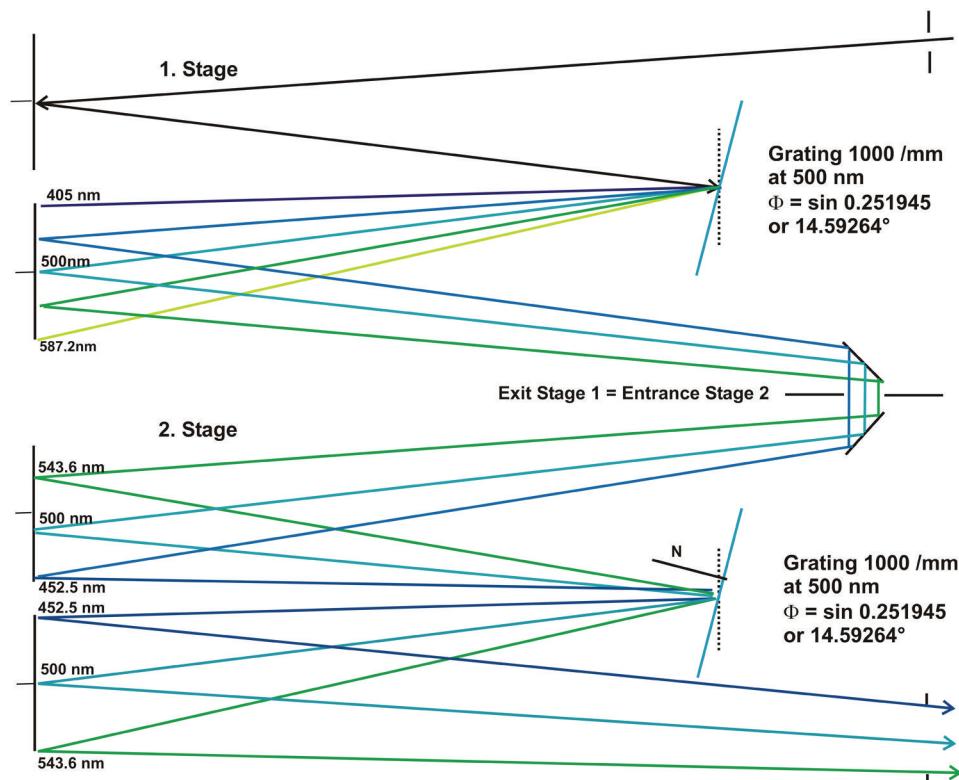


Figure 4.23 Basic beam travel of a classical, additive, double Czerny–Turner or Ebert–Fastie.

used with Eq. (2.1) at 500 nm ($\sin \alpha + \sin \beta = 0.5$). The focal length is assumed to be 500 mm, and the width of the mirrors is 80 mm; the grating will be wide enough to allow full illumination, and thus $\varepsilon_1 = \varepsilon_2 = 7.125$ deg. The mirror-grating distance will be 400 mm. The gratings are set to $\phi = 14.59$ deg, or 500 nm in first order; they turn towards the entrance. The system moves the two gratings parallel by the same screw. The focus mirror of stage 1 will receive an interval of 182.2 nm, or 2.27 nm/mm. The colors of the sketch roughly fit the wavelengths involved. Because stage 2 doubles the wavelength spread, the center wavelength interval is directed through the common intermediate aperture, which is the interval blue to green.

To get 500 nm from the center of focus mirror 2, angle $\beta = 21.717$ deg. To create the sum ($\sin \alpha + \sin \beta = 0.5$, $\alpha = 7.47$ deg. On the inner side of the focus mirror, there is the wavelength 452.5 nm at β (blue) = 16.024 deg, which means that β_{blue} needs to be 10.166 deg. On the other side of the mirror, the green 543.6 nm is at $\beta_{green} = 4.899$ deg.

Everything is fine, but as the graph shows, collimator 2 will be illuminated out of the center. Because the two gratings are fixed to the same drive, the deviating mirror after the intermediate aperture (center slit) will be adjusted to correct that. Because the “mismatch” is a constant angle for all wavelengths and all gratings, that is an acceptable solution.

4.3.3.2 Subtractive setup

In the additive system, the intermediate aperture only needs to transfer the part of the spectrum that will leave stage 2 at the end. Normally, that is the center part of the spectral interval created by stage 1. In subtractive mode, the interval passing the intermediate slit represents the wavelength interval to be united in the output.

If, in Fig. 4.23, the grating of stage 2 turns towards the output, all wavelengths entering would be multiplied with the λ factor -1 , which inverts the order used. Consequently, the dispersion would be compensated, as well. It would be the perfect neutralization of the separation: at every geometric position in the output, the light would be a mix of the interval entering stage 2. The grating dispersion and the intermediate aperture define the spectral bandwidth. The entrance and exit slit (or aperture) define the area. This arrangement would be the ideal subtractive solution. Figure 4.24 presents a version that fits this model as closely as possible.

The simplified illustration shows a subtractive, parallel tandem spectrometer that is essentially the same as that used to discuss additive ray tracing. Because stage 2 inverts the dispersion, the light in the output will be the mixture of all wavelengths entering stage 2. In the case shown, that is blue to yellow. To reach zero dispersion in the output, the product of all sine functions affecting a wavelength within the interval must have the same value ($\sin 0.5$, for instance). To provide that, a beam inversion is required between stages 1 and 2, performed here by the external elliptical mirrors. The spectral

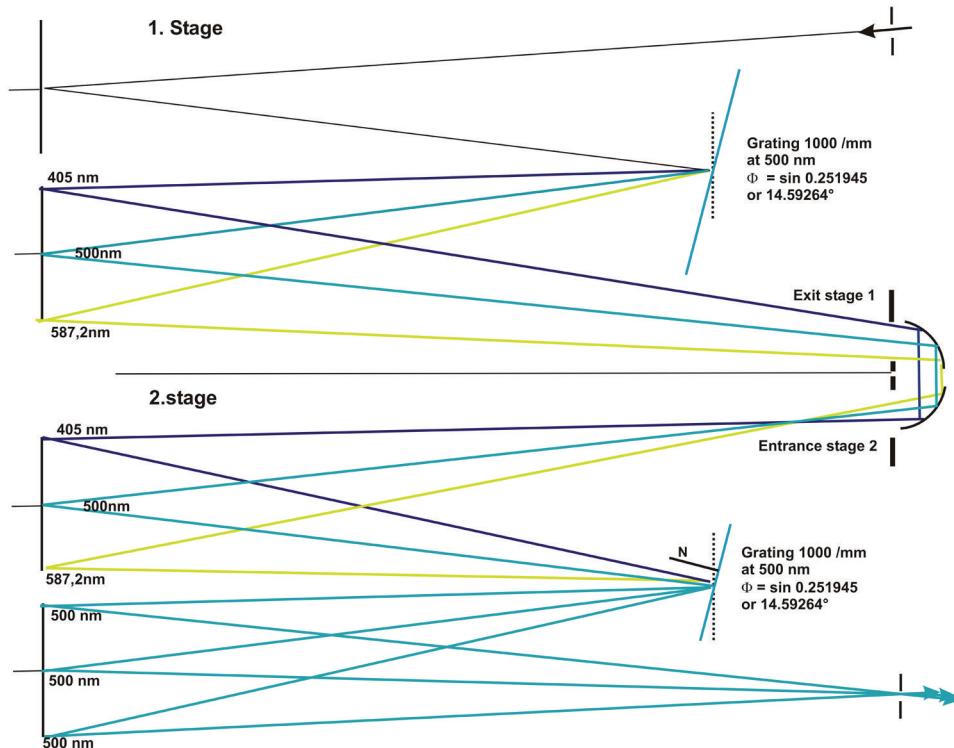


Figure 4.24 Principal beam travel of a subtractive, classical Czerny–Turner double spectrometer.

interval then leaves the output of stage 1 through a slit and enters stage 2 after the inversion through one more slit. It naturally requires a highly stable, well-calculated, and well-adjusted system to work properly. To create 500 nm (turquoise) in the center of focusing mirror 2, the required angles are $\alpha = 7.4676$ deg and $\beta = 21.717$ deg. On the inner side of the mirror, there is the wavelength 405 nm at $\beta_{blue} = 16.024$ deg. For full dispersion inversion, the sum of both sine functions also must be 0.5; thus, $\alpha_{blue} = 7.4676$ deg. On the yellow side, at 587.2 nm, $\beta_{yellow} = 27.272$ deg, which requires $\alpha_{yellow} = 2.395$ deg. Figure 4.24 shows that for all wavelengths, the inclusion angle between a beam entering the grating and then leaving it is the same for all rays. In the output, all colors are then effectively overlaid, creating the mixed-wavelength band. The instrument described in Figs. 4.20, 4.23, and 4.24 is useful for both additive and subtractive modes. The beam-diverting mirrors direct the beam travel between stages to the required path.

4.3.3.3 Modern off-axis double spectrometers

So long as some precautions are not ignored, modern off-axis double spectrometers can be configured for additive or subtractive work. Figure 4.25 shows that upon first dispersion, the light will either be deviated to stage 2 or

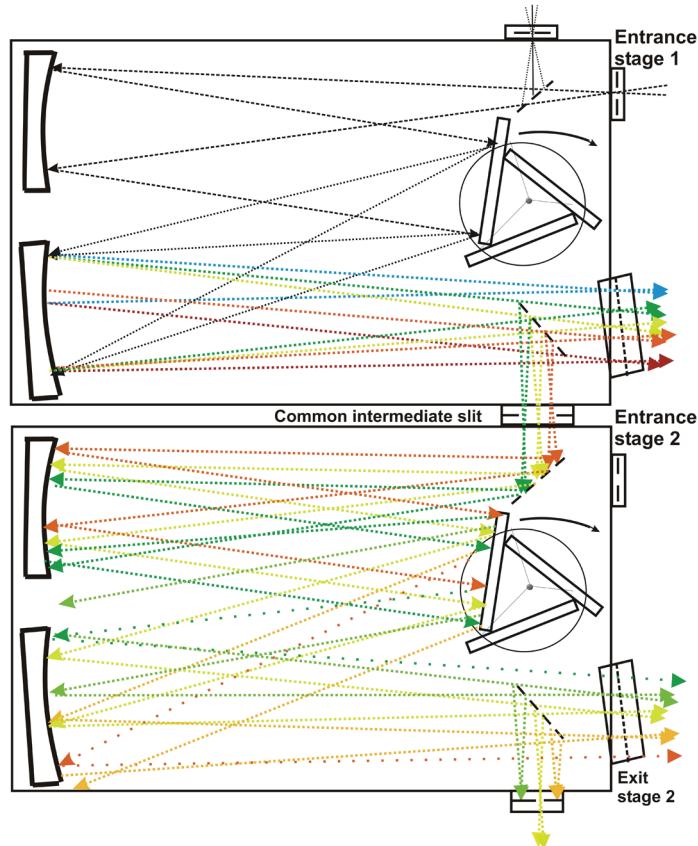


Figure 4.25 Flexible Czerny–Turner double spectrometer with a triple off-axis grating turret, additive operation.

to an area detector. In additive dispersion, the grating of stage 2 turns parallel to that of stage 1 (all rules described in Section 4.3.3.1 apply). To provide subtractive dispersion, the second stage is inverted. Note that a design like that shown in Fig. 4.26 will automatically lead to asymmetries and imperfect subtraction because the gratings do not turn in the front surface. These flaws are indicated in the graph by the imperfect colors in the output.

4.3.3.3.1 Methods to achieve subtractive dispersion

The subtraction effect can be created in several ways. Figure 4.24 shows the external wavelength inversion, whereas Fig. 4.26 demonstrates inversion by inverted grating rotation. Furthermore, the system of Fig. 4.26 is based on a triple off-axis grating turret. The diverting mirror in the exit of stage 1 cuts out a limited spectral range (in the example, green to orange) and sends it through the intermediate aperture. It is assumed that 20 nm (570–590 nm) will reach stage 2. Because the interval originates from the center part of stage 1's focus mirror, it will be rather homogenous in intensity. The different wavelengths

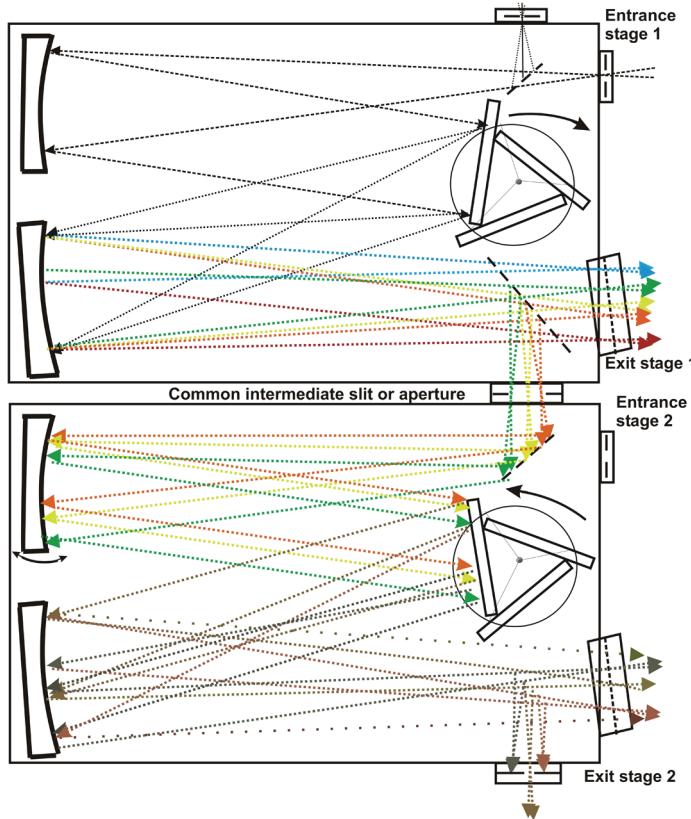


Figure 4.26 Asymmetric, subtractive double spectrometer supported by off-axis gratings.

leave stage 1 under different angles and at different geometric positions. The differences are transferred to the collimator and grating of stage 2. If the grating of stage 2 would turn towards the exit, as described after Fig. 4.23, the result would be perfect subtraction because of the perfect optical symmetry. However, the grating drive is a single shaft, inhibiting the different rotation. In Fig. 4.26, the drive is independent, but the optical system becomes asymmetric as soon as the grating turns towards the exit.

The following assumptions are made: The system will be used for additive and subtractive work; $f = 300 \text{ mm}$; the width of the mirrors is 75 mm ($f/4$); and the median grating-mirror distance is 240 mm (80% of f). The front surface of the gratings rotate at a radius of 40 mm , the gratings are 1800 mm^{-1} , $\varepsilon_1 = \varepsilon_2 = 12 \text{ deg}$, and the test wavelength is 580 nm (yellow).

Using Eqs. (2.17)–(2.20), the dispersion of stage 1 is found to be 1.326 nm/mm . To transfer 20 nm , the intermediate aperture width needs to be 15 mm wide. According to Fig. 4.4 in Section 4.1.1, the intensity drops towards the outer ranges of the field. The grating turret is placed perfectly in the center axis of the spectrometer. Because the grating rotates on a radius, when set to 580 nm (or 8.5 deg), it moves 8 mm towards the entrance. The working angles

became $\alpha = 20.25$ deg ($\sin \alpha = 0.3461$) and $\beta = 44.25$ deg ($\sin \beta = 0.6979$). The angular path between both angles leads to $\delta = 24$ deg. The desired wavelength travels at $\beta = 42$ deg (related to N). The two extreme wavelengths of the transmitted band (570 and 590 nm) shine up under the angles $\beta = 49$ deg (“red side”) and 36 deg (“blue side”). These values are then used in Eqs. (2.1) and (2.5).

The following differential dispersion is found for the three colors. The reference is 580 nm ($\cos \beta = 0.7162$, or 44.257 deg): At 570 nm ($\cos \beta = 0.7333$, or 42.834 deg), the differential $\delta\beta$ equals $\arccos 0.0171/10$ nm. At 590 nm ($\cos \beta = 0.69812$, or 45.715 deg), the differential $\delta\beta$ is $\arccos 0.01798/10$ nm. The result is a variation in dispersion of ~5% within the intermediate aperture.

Stage 2 is equivalent to stage 1, but the grating rotates in the inverse direction in order to compensate for the dispersion of stage 1. If the spectral orders of stage 1 sustain a positive sign, the sign will become negative in stage 2. Theoretically, a perfect balance between both stages would take place if the same values were found again. Due to the inverted grating rotation, the value of angle α in stage 1 becomes the value of β , and vice versa. Because the rotation introduces small asymmetries, a slightly different angle β will be found between the grating and focusing mirror. The wavelengths of the band transmitted (570–590 nm) then have the angles $\beta = 29$ deg (“red”), 20 deg in the middle, and 12 deg (“blue”).

Entering those angles into Eq. (2.5) produces new results. The reference is 580 nm ($\cos \beta = 0.9382$ or 20.25 deg). At 570 nm ($\cos \beta = 0.94463$, or 19.155 deg), the differential of $\delta\beta$ is $\arccos 0.00643/10$ nm, and at 590 nm ($\cos \beta = 0.93135$, or 21.353 deg), the differential of $\delta\beta$ is $\arccos 0.00684/10$ nm. Compare the values with those presented for stage 1 earlier. Even if the median dispersion is the same, the differentials are different. This difference means that the band transmitted is intact but not perfectly subtracted; residual errors will remain, as Fig. 4.26 illustrates in the output. On one side, the lower wavelengths will be more strongly represented compared to the other end, and vice versa, while in the center they will be well mixed. A system like this is called semi-subtractive.

The same effect will be found in fully symmetric systems, too, with the grating rotating in the perfect position. However, due to the missing off-axis shifts, the impact is reasonably less. The different entrance angles lead to small angular shifts at the grating for the different wavelengths, causing a small imperfection. Of course, the error will be much smaller, but the wider the spectral band and the larger the field is, the larger the error. The upshot is that there are ways to improve. One option for inverse rotating systems involves readjusting (horizontally tilting) the collimating and/or focusing mirror with a flip of the grating, which will improve the illuminated area of the grating and can correct the angles within certain limits. In a system that is used in both additive and subtractive modes, the rotation of one grating changes. A motorized mirror adjustment could be the solution. Because different calibration curves must be used for the different modes anyway, all parameters can be stored and recalled on demand. No further spectroscopic

disadvantages need be dealt with. In Fig. 4.26, that option is marked by darts at the collimator. In a modern system, the realization is mainly a question of electronics and software.

4.3.3.4 Mechanical filtering in double spectrometers

Figures 4.23–4.26 illustrate the focal plane in the area of the intermediate slit where the different wavelengths pass in a parallel manner. Each color has its own focal point, which can be used to place a mechanical filter. Imagine a strong but unwanted signal passing through the aperture, as with a Raman measurement, which records Stokes and anti-Stokes signals simultaneously. The strong laser line can be cut out by a thin, needle-shaped, vertical beam stop in the focal plane. Equipped with a fine positioning adjustment, the beam stop can be moved to the best position or out of the spectral field. Naturally, after the adjustment, no further movement of the grating in front of the beam stop is allowed. A certain wavelength or a small interval will be strongly reduced—at the expense of increased stray light from the extra mechanical parts, which can create diffusion problems. It is necessary to perform test runs and calculate the advantages versus disadvantages when using such an option. For scanning systems, this option is not useful; they suffer much less from stray light, due to the narrower slits, which allows data gathering close to the disturbing signal, and a shutter can close while scanning over the critical spectral zone.

4.3.4 Various configurations of flexible double spectrometers

The design portrayed in Fig. 4.27 is made possible by separated-grating control systems and provides maximum flexibility and efficiency. Utilization of the front entrance and exit means that only six reflections occur in total—that provides transfer efficiencies up to 70% and dispersion efficiencies above 50%.

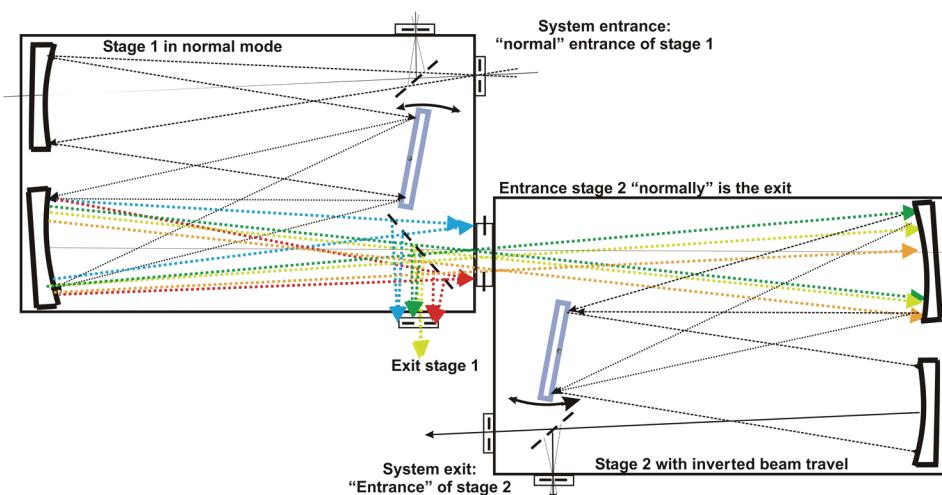


Figure 4.27 A front–front coupled double configuration for additive and subtractive mode.

Indeed, special coatings are required to achieve this, working only over a limited range, and the high dispersion efficiency only appears near blaze condition. But for Raman or other narrow-bandwidth applications, these values are realistic. In Fig. 4.27, the dispersed light from stage 1 enters stage 2 through the “output” of 2, and it finally leaves through its “entrance.” As noted earlier, this arrangement is no problem for monochromator systems. The figure displays the subtractive mode because the grating of stage 2 has turned towards the actual exit, whereas the grating of stage 1 has turned to the entrance. That kind of setup even allows a symmetric transfer function with asymmetric spectrometer configurations. In a cleverly designed system, both gratings can be turned towards the entrance or exit independently to allow all kinds of double configurations. Because that arrangement requires several sets of wavelength calibration parameters, it is more a matter of electronics and software, which can also include the motorized optimization of the mirrors. The result is a highly efficient and flexible system that can provide several modes of operation:

- Both gratings turn towards the actual entrance or actual exit of both stages. This behavior provides additive dispersion; the side output of stage 1 and both exits of stage 2 can be equipped with all kinds of detectors.
- The grating of stage 2 turns inversely to stage 1, which provides subtractive dispersion. An area detector only makes sense at the side output of stage 1. Both exits of stage 2 can be equipped with single detectors or be used for homogenous area illumination. They can also illuminate the entrance slit of a third stage.
- Additive and subtractive double spectrometers can be based on Ebert–Fastie, Czerny–Turner, normal-incidence, and Seya–Namioka concepts. Side–side, front–front, and side–front configurations have been realized for all kinds. The highest flexibility is provided by Czerny–Turner, whereas Ebert–Fastie delivers the best performance.

4.3.5 General performance data of double spectrometers versus similar single-stage systems

The following parameters are listed in Table 4.2.

- **Resolution:** R_p is defined by the focal length, the grating parameters, and the minimum slit or pixel width. In double spectrometers, the synchronicity of the two stages and the transfer of the entrance image to the exit image are the limiting factors. A single spectrometer with a focal length equaling the sum of the two stages of a double will always have the better practical resolution.
- **Luminosity:** L is given by the gratings and the slit areas in relation to the focal length. Using the same parts, an additive double will have twice the aperture because the slits are twice as wide at the same bandwidth. However, it also has twice as much reflection loss. In total, L can be similar for both versions.

- **Contrast:** C and stray light reduction are the overwhelming and decision-making advantages of the double configuration. The second stage only receives a pre-selected and pre-cleaned spectral interval that is processed again, leading to an exponential increase in contrast.
- **Stability:** Although single-stage instruments will suffer from instabilities and drift, those effects can be fatal for a double if the two stages drift apart. The light beam can be blocked in the worst cases. Single-stage instruments tolerate much stronger instabilities.
- **Working Range:** Because all losses and polarization phenomena occur twice in doubles, the useful working range is narrower compared to similar single-stage units. In most cases, grating changes need to be done more often, and the selection of gratings is more difficult.

4.3.6 Triple-stage spectrometers

To increase dispersion and/or stray light reduction further, a third stage can be added. In an additive three-stage spectrometer, the dispersion equation changes to

$$RD-TA = \lambda / [(2f_1 \times \tan \phi_1) + \lambda / (2f_2 \times \tan \phi_2) + \lambda / (2f_3 \times \tan \phi_3)]. \quad (4.4)$$

In a subtractive system, stage 2 compensates the dispersion of stage 1, and stage 3 creates the final dispersion of the whole system:

$$RD-TS = \lambda / [(2f_1 \times \tan \phi_1) - \lambda / (2f_2 \times \tan \phi_2) + \lambda / (2f_3 \times \tan \phi_3)]. \quad (4.5)$$

The additive setup of three stages is almost only found in scanning monochromator systems. The advantage is again the exponential contrast improvement compared to the already good contrast of a double system. Triple monochromators can acquire valid data in very close vicinity to the disturbing source, such as a laser. For example, a three-stage 750-mm monochromator with gratings of 1800 mm^{-1} will create a dispersion of $\sim 0.2 \text{ nm/mm}$, or $7.7 \text{ cm}^{-1}/\text{mm}$ in the green range. With slit widths of $100 \mu\text{m}$ (resulting in a resolution of perhaps 10 pm), it will provide good data down to < 0.4 wavenumbers (which is 5 bandwidths apart from the laser) and have a stray light suppression of $>> 10^{10}$. In order to get even closer, either the slits must be tuned narrower or else increased stray light must be tolerated. To perform measurements extremely close to the disturbing wavelength requires special instruments that provide valid data down to $\sim 1 \text{ pm}$ or 0.1 cm^{-1} relative to the source.

Monochromator data acquisition with small bandwidths produces the most unambiguous data of all available methods. Of course, they need long acquisition times, especially at low light levels. That, in turn, explains the popularity of parallel detection for that family of measurements. With the availability and improvement of efficient array and CCD detectors, a new generation of triple spectrometers was developed, effectively combining a subtractive pre-stage with a high-resolution end stage. If used as an additive

triple with parallel detection, all slits except the entrance need to be wide open, allowing lots of unwanted light to reach the area detector—thus, the concept of running the first two stages in subtractive pre-stage mode, which strongly eliminates stray light from the entrance of stage three, is the most popular choice. From the dispersion and resolution perspective, only stage three, the spectrograph, is important. A multi-grating system is often used for enhanced flexibility. Also, the focal length of stage three is often longer than that of the two pre-stages; they define the interval transferred into stage three and dominate the stray light rejection. It is important to align the acceptance and transfer angles of the stages. The transmission efficiency of those systems can vary strongly with wavelength, which requires serious consideration during the time of definition. A highly efficient yet flexible setup is shown in Fig. 4.28.

The system only needs nine reflections in the subtractive mode of a triple-stage spectrograph. In a typical case and close to blaze condition, T will be $\sim 15\%$. That is the case if the system must provide a wide useful spectral range and has no specially coated mirrors; the range of 250–1250 nm (five spectral orders) is no problem with a set of three gratings installed. If optimized to a single spectral order, i.e., 400–800 nm, and optimized coatings, T can exceed 30%. The subtractive pre-stage might only need one grating to cover the limited range. The considered system offers optimum flexibility.

The third stage works equally well as a spectrograph and a monochromator. It features an extra fiber optic entrance at the side input proposed for a

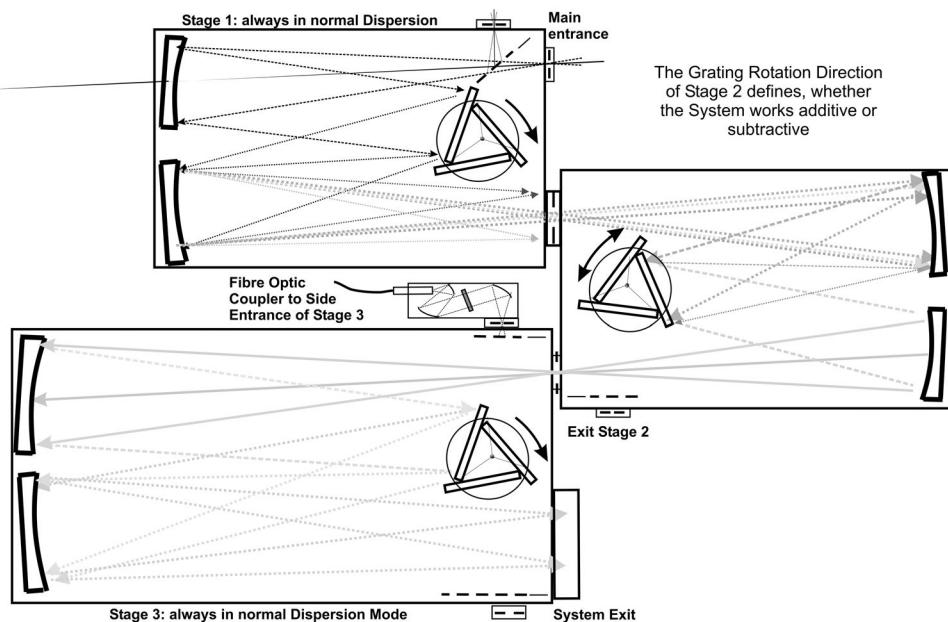


Figure 4.28 Triple-stage spectrometer for single-, double-, and triple-stage use in additive and subtractive operation.

Raman notch filter. Using that entrance, stage 3 will behave and provide the same $T (>0.5)$ as a single-stage setup. The second stage's exit is reserved for a single-wavelength detector (photon counter) and scanning mode. The first two stages work in subtractive and additive mode. In subtractive mode, the second grating turns clockwise towards the exit of stage 2. It will shed a selected wavelength interval at the small entrance slit of stage 3, which also represents the exit slit of stage 2. If the grating of stage 2 rotates counter-clockwise, the dispersion becomes additive, and stage 3 receives a highly pre-dispersed signal. If the combined stages 1 and 2, and stage 3 are driven by separate PCs, they can run completely different experiments.

It would exceed the scope of this book to go into further detail about multi-stage spectrometers. However, this volume's companion text regarding applications⁴ addresses the advantages of multi-stage systems and demonstrates why they are often recommended, sometimes as the only solution.

4.4 Echelle Spectrometers

As described in Section 3.3, an Echelle grating features several exclusive parameters:

- The short side of the structure is illuminated instead of the long edges. The angle between the edges is always 90 deg.
- To illuminate the short edge generally requires larger working angles. Larger angles lead to higher dispersion, which in turn leads to higher resolution. Due to the two right angles, the working angle ϕ cannot be smaller than 45 deg. Blaze angles up to 80 deg are produced. Some ruled gratings with angles <45 deg can be obtained featuring similar parameters as Echelles; they are called Echellette gratings.
- Because of the short edge, the useful angular range of an Echelle is smaller than the angular range of a standard grating. This is due to shadowing effects and quickly decreasing surface efficiencies with angle variation. Echelle instruments must be designed to be very slim, which is advantageous for Ebert–Fastie systems.
- Echelle gratings generally have low line densities and offer very high surface quality. These traits apply to the substrate and the micro structures (edges), as well. The benefit Echelles provide is a rather homogenous efficiency behavior over a wide range of spectral orders. That is their major difference with holographic gratings, which, in practice, only work in one spectral order. Some classically ruled gratings also work over several orders (the data sheets usually indicate this). Echelle spectrometers are useful between <200 nm to >10,000 nm with little variation in efficiency if the grating is kept near the blaze angle (see Fig. 3.3).
- If the entrance spectrum is wider than one order, Echelle spectrometers require blocking towards lower and upper orders, as well.

4.4.1 Echelle monochromators and 1D spectrographs

A spectrometer with internal angles ε_1 of <10 deg is generally suited for the use of Echelle gratings. Consider an Ebert–Fastie with a 1.2-m focal length, featuring an aperture of $f/10$. The mirrors will have an illuminated height and width of 120 mm each. Following the Fastie rules and allowing a light trap of 30-mm width in the center of the Ebert mirror, the mirror will have a total width of 270 mm and a height of 120 mm. The distance between the mirror and the grating will be ~ 960 mm. Assuming perfect illumination of both the mirror and grating, both ε are 4.47 deg. A grating of 158 mm^{-1} at 63-deg blaze ($\cos = 0.454$) is then chosen. To provide an illuminated width of 120 mm, the grating must be at least 265 mm wide, which is available. Calculation of the first-order wavelength is conducted using the Ebert equation [Eq. (2.8)]: $m\lambda = k \times (2 \sin \alpha) \times \cos \varepsilon$, and one finds that in the first order $\lambda = 11.24 \mu\text{m}$. If one were to analyze, for example, the laser line of He-Ne at 632.8 nm, it would be almost exactly at blaze in the 18th order. The dispersion is then calculated by Eq. (2.20): $RD = \lambda/(2f \times \tan \phi) = 0.135 \text{ nm/mm}$.

The Echelle will be a rather expensive grating, but what are the alternatives? A standard grating of 2400 mm^{-1} would have a working angle of $\phi = 49.75$ deg. To realize $f/10$, it needs a width of 185 mm, which is not available for standard gratings. Assuming that one can be obtained, the first-order dispersion will be 0.223 nm/mm, producing a resolution that is worse by a factor of 1.65. The Echelle will work rather uniformly between 200 nm and 11 μm , and will create only small polarization effects. Another benefit is that its dispersion and resolution will increase towards the UV. If the resolution is still not satisfactory, Echelles with higher line densities and/or higher blaze angles can be chosen, as shown in Fig. 4.29. A standard grating, either holographic or ruled, can be optimized to 400 nm and work from 300–700 nm with remarkable efficiency variation and strong polarization effects. The resolution will increase towards the NIR, which is most often not an advantage. At 633 nm, the free spectral range (FSB) in the first order is ~ 315 nm. For the Echelle chosen, it is [by Eq. (2.3)]: $\lambda_2 = \lambda_1 + (\lambda_1/m) = 634 \text{ nm}/(18 \times 2) = 17.6 \text{ nm}$.

Consequently, in applications with spectral information wider than the actual order, filtering higher and lower orders is required. In addition, the grating control system needs improved flexibility. Scanning ranges wider than one order includes changing orders and repeatedly scanning with different parameters to achieve a wide spectrum. The grating will only be moved in a narrow angular range while the actual order selection occurs externally. The best pre-selector is a prism monochromator in front of the main unit. A wide range of prism materials exists, and most of them have the physical advantage of increasing the system dispersion towards the UV. With modern stepper-motor controllers and microprocessor systems, it is no challenge to synchronize two axes and offer flexible control parameters.

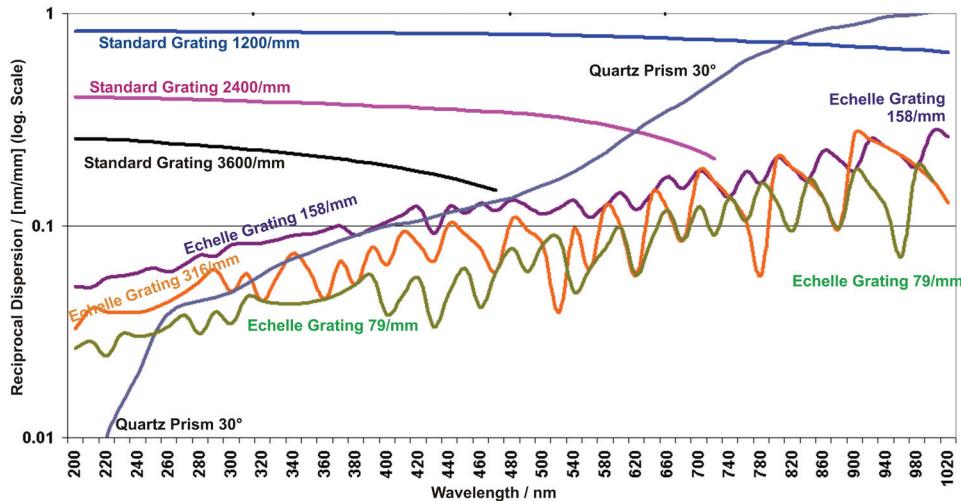


Figure 4.29 Dispersion behavior of different dispersers, normalized to a 1-m focal length, for high-resolution measurements. The wavelength axis is linear, and the dispersion axis is logarithmic.

The chart in Fig. 4.29 compares three types of dispersers:

1. Three standard gratings (ruled or holographic) in the first order. The line densities from top are as follows: 1200 mm^{-1} (blue, reaches up to 1500 nm), 2400 mm^{-1} (pink, reaches up to 740 nm), and 3600 mm^{-1} (black, reaches up to 450 nm).
2. Echelle gratings of 158 mm^{-1} , blaze 63 deg (violet, reaches up to 12 μm); 316 mm^{-1} , blaze 70 deg (orange, reaches up to 6 μm); and 79 mm^{-1} , blaze 76 deg (green, reaches up to 24 μm).
3. A quartz prism (grey curve) with base angle $A/2 = 30$ deg, in reflection mode (useful for 160–1300 nm). Below 250 nm, the prism's dispersion will change quickly with deflection angle d and illumination angle i_1 .

Reviewing the data near 633 nm shows that the grating of 1200 mm^{-1} provides an RD of 0.767 nm/mm, the grating of 2400 mm^{-1} creates an RD of 0.265 nm/mm (at a working angle already at 50 deg), and the 3600-mm^{-1} grating is out of range. The Echelle of 316 mm^{-1} in 18th order provides 0.143 nm/mm, that of 158 mm^{-1} in tenth order provides 0.11 nm/mm, and that of 79 mm^{-1} in 38th order provides 0.09 nm/mm. The FSB of the 79-mm^{-1} model is 8.4 nm. The Echelle gratings will only be used in a narrow angular range. The frequent switching of orders leads to a modulation of the efficiency curve. The 30-deg quartz prism in a 1-m spectrometer would have a RD of ~0.3 nm/mm at 633 nm.

4.4.2 High-resolution Echelle spectrometer designed as a monochromator and 1D spectrograph

Based on the calculated data, an Echelle spectrometer can be safely configured, as shown in Fig. 4.30. After the grating for the main stage is selected, the prism

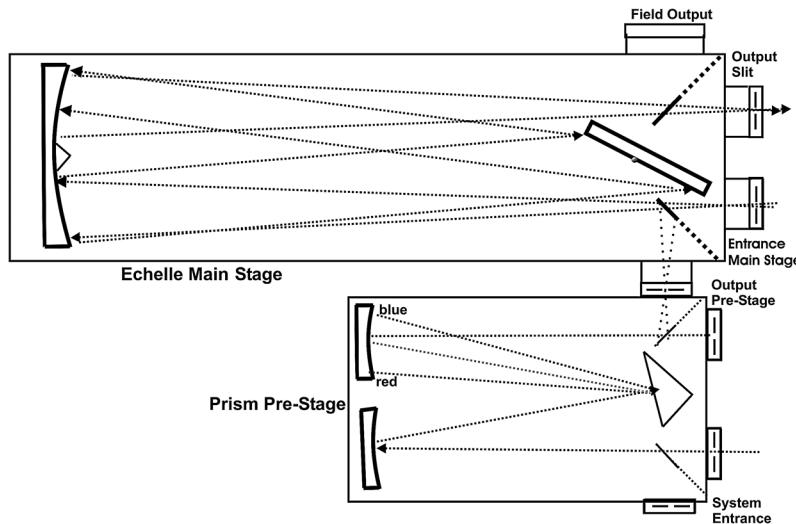


Figure 4.30 Echelle spectrometer for wide wavelength ranges.

pre-monochromator needs to provide the filtering required. Its definition in a monochromator system is easy; only two parameters need to fit:

- The bandwidth delivered by the prism must, at any measured wavelength and slit width, not be smaller than the bandwidth produced by the Echelle stage. This is ensured if the RD number of the prism stage is always larger than that of the main stage.
- The bandwidth entering the Echelle stage must always be within one spectral order. If the bandwidth would be wider than one order at fully opened slits, order overlay can occur. Eventually, the entrance slit width of the main stage must be limited to avoid that scenario.

For the calculation in spectrograph mode, i.e., with a CCD detector, keep in mind that even at the minimum entrance-slit width (for optimal resolution), the full spectral interval captured by the CCD needs to enter the main stage. Thus, the definition of a prism–Echelle combination is an iterative process. The following simulates a typical step.

The main unit must work with both a single-point detector and a CCD. The useful interval will be the typical working range of a CCD: 200–1000 nm. Because a high-resolution system is desired, an iteration with $f = 1.2$ m is chosen; the CCD detector is 25.6 mm wide and 12.3 mm high, presenting 1024×512 pixels of $25 \times 25 \mu\text{m}$. An Echelle of 158 mm^{-1} is initially selected at 63-deg blaze. The size of the ruled area is 200 mm \times 100 mm. At $63 \text{ deg} = \cos 0.454$, the illuminated width becomes 90 mm, which will also be the illuminated height. The resulting aperture will be $f/13.3$. A light trap of 10 mm width is placed in the center of the Ebert mirror, leaving both halves of the mirrors 50 mm away from center. That, in turn, leads to $\tan \epsilon = 0.052$, or $\epsilon = 3 \text{ deg}$ (rounded). Also by

choice, a prism pre-stage of $f = 320$ mm is used, equipped with a quartz prism. To ensure full throughput at all wavelengths, a slightly wider aperture of $f/12$ is allowed. The illuminated prism area needs to be at least 27 mm wide and high; the example is 30 mm. To start calculations, the internal angles of the pre-monochromator are set to $\delta = 30$ deg, leading to $d = 30$ deg, too, starting at $i_1 = 15$ deg. The distance of the mirrors from the center is 33.5 mm.

Parameters at $\lambda = 1000$ nm

The RD of the Echelle stage at $\lambda = 1 \mu\text{m}$ swings $\sim 0.3 \text{ nm/mm}$ in the 11th order, and the FSB is 45 nm. The CCD will receive an interval of $\sim 7.5 \text{ nm}$. The optimal CCD resolution requires a slit width of 50 μm , resulting in a bandwidth of 0.035 nm at 1000 nm. Using the slit output, the resolution will be $\sim 3 \text{ pm}$. At 1000 nm, the prism pre-stage shall deliver a bandwidth between 7.5 nm and 45 nm, into a slit width of 50 μm . The exit slit of the pre-stage and the entrance slit of the main unit are identical. With $A/2 = 30$ deg of the reflection prism, at $f = 0.32 \text{ m}$, the RD is $\sim 700 \text{ nm/mm}$. Within a 50- μm slit width, the bandwidth is 35 nm, which is very good because it is more than the CCD interval but less than one order. The limiting element in that design is the maximum allowed slit width of 65 μm ; otherwise, order overlay occurs.

Parameters at $\lambda = 500$ nm

The RD of the 1.2-m Echelle is 0.092 nm/mm in the 23th order, and the FSB is 10.9 nm. The CCD is illuminated by 2.3 nm. The prism in the 32-cm pre-stage provides 150 nm/mm. Within a slit width of 50 μm , the bandwidth is 7.5 nm, which fits very well, too. The slits can be opened up to 73 μm with no problems. The monochromator resolution will be 2 pm, and with a CCD it will be 11 pm.

Parameters at $\lambda = 300$ nm

The RD of the 1.2-m Echelle is 0.06 nm/mm in the 37th order, and the FSB is 4 nm. The CCD receives 1.54 nm. The prism will deliver 30 nm/mm, or 1.5 nm, into a 50- μm slit width. The entrance slit of the main stage needs to be opened to 52 μm to allow the 1.54 nm to enter. Up to 2 μm more than the optimal 50 μm is no real problem. The slits can be opened to 130 μm without penalty. In monochromator mode, the resolution will be 0.6 pm, and with a CCD it will be $\sim 3.8 \text{ pm}$. This is at the limit but still within acceptable parameters.

Parameters at $\lambda = 200$ nm

At this point, the 1.2-m Echelle is in the 56th order and provides an RD of 0.043 nm/mm; the FSB is $\sim 1.8 \text{ nm}$. The prism spectrometer, as defined so far, will have an RD of 3 nm/mm, leading to a maximum slit width of 0.6 mm to avoid order overlay. The CCD will see an interval of 1.075 nm. A 50- μm slit will transfer only 0.15 nm.

To provide 1.075 nm, a slit width of 360 μm would be required. In monochromator mode, the resolution will be ~ 0.4 pm. With a CCD and a 50- μm entrance slit, the resolution is ~ 2.6 pm, increasing to 11 pm with a 360- μm slit. An obvious dilemma appears: one can achieve very good performance with a single-point detector, but with a CCD the resolution will worsen towards the deep UV because the slit needs to be ever wider. The reason for this behavior is the extreme increase in prism dispersion towards the UV.

Regardless, once the first calculation sequence is done, it is time to review all parameters and requirements. One issue is the question of whether parallel detection of only 1.075 nm at 200 nm makes sense at all. Assuming that the spectral interval can be widened at all wavelengths at the expense of resolution, a prism can be selected with a smaller angle $A/2$, the focal length of the pre-stage can be shortened. Alternatively, because dispersion depends on the working angle of the grating, an Echelle with a lower blaze angle will also reduce dispersion. Echelle-compatible systems can be built down to focal lengths of ~ 600 mm. After this decision is made, the next iteration can take place.

4.4.2.1 Echelle aberrations

Assuming that the system defined so far is acceptable, one must look up the aberrations. For the prism spectrometer, this involves Eq. (2.26), $O_1 = 1.3 \mu\text{m}$, and Eq. (2.27), $E = 1.7 \mu\text{m}$, presuming straight slits of 12.3-mm height. The horizontal aberration of Eq. (2.28) will be $H = 9.6 \mu\text{m}$. The vertical slit image is 4.7 mm, but it has no impact on dispersion or the resolution of the main unit. All numbers are the worst case and are related to the maximum slit width of 5 mm; much better behavior can be expected during real operation.

Calculating the main unit as a monochromator with slits of 50- μm width and 5-mm height (using straight slits), it is found that $O_1 = 3 \mu\text{m}$ and $E = 3.4 \mu\text{m}$. With curved slits, $E = 4.3 \mu\text{m}$. $H = 9 \mu\text{m}$ with straight slits, but only 3.4 μm with curved slits and slits fully opened. In spectrograph mode and with a detector length of 25.6 mm illuminated to 12.3-mm height (with a straight slit of 50- μm width), $O_1 = 1.5 \mu\text{m}$, $E = 1.75 \mu\text{m}$; and $H = 5 \mu\text{m}$. All numbers create no problem at operation, because all are well below the pixel size of 25 μm .

4.4.2.2 Thermal drift assuming an aluminum chassis

The drift of the prism stage is found by Eq. (2.31): $dx = K \times 2f \times dT$. Applying the data $K(\text{Aluminum}) = 2.2 \times 10^{-5}/^\circ\text{C}$, $2f = 640$ mm, and $dT = 1^\circ\text{C}$ produces $dx = 14 \mu\text{m}$.

The defocusing factor, according to Eq. (2.31), is $dy = dx/n$, $dy = 0.014 \text{ mm}/12 = 1.17 \mu\text{m}$, a good value. The effect of the mirror substrate is then calculated, which works in the counter-direction, $K(\text{Glass}) = 0.86 \times 10^{-5}/^\circ\text{C}$, and both mirror substrates are 10-mm thick. The prism will not be taken into

account because it moves. Thus, $dx_{Optics} = 0.86 \times 10^{-5}/^\circ\text{C} \times 2 \times 10 \text{ mm}/1 \text{ }^\circ\text{C} = 0.17 \mu\text{m}$, which can be neglected. All values found are normalized to 1 °C and thus must be multiplied by the expected thermal change. Finally, the results are compared with the slit width to understand the impact.

The drift of the main unit is $dx = K * 2f * dT$, where $K(\text{Al}) = 2.2 \times 10^{-5}/^\circ\text{C}$, $2f = 2400 \text{ mm}$, and $dT = 1 \text{ }^\circ\text{C}$; the result is $dx = 52.8 \mu\text{m}$. The defocusing factor is $dy = dx/n$, where $dy = 0.0528 \text{ mm}/13.3 = 4 \mu\text{m}$. Recall the compensation by glass: $K(\text{Glass}) = 0.86 \times 10^{-5}/^\circ\text{C}$. The Ebert mirror will be 25 mm thick. Because the grating is under 63 deg, it is omitted, and thus $dx_{Optics} = 0.86 \times 10^{-5}/^\circ\text{C} \times 2 \times 25 \text{ mm}/1 \text{ }^\circ\text{C} = 0.43 \mu\text{m}$. The resulting defocusing impact is $52.4 \mu\text{m}/^\circ\text{C}$, which means that the drift can become fatal at small slit widths. Consequently, unless the environment is extremely stable, the thermal situation of the system must be stabilized (discussed at the end of the next section).

4.4.3 Two-dimensional Echelle spectrometer for the parallel recovery of wide wavelength ranges at high resolution

As demonstrated earlier, the combination of a prism and an Echelle grating is useful for wide wavelength ranges, providing flexible and order-free parameters. (A 2D detector can be used to review it.) The prism is placed between the grating and the detector, and the dispersion function of the prism is organized so that it separates the spectral orders in a vertical manner. These orders are called spectral intervals, arriving vertically on top of each other in “stripes.” In the “horizontal” orientation, the side-by-side dispersion of wavelengths within one order appears. The grating dispersion of the instrument must behave such that stripe wavelength overlay does not occur anywhere; otherwise, they cannot be separated. On the other hand, the wavelengths of subsequent orders, lying on top of each other, must not have any wavelength gap in order to allow the full recombination of data. Thus, one stripe must contain at least one full order but less than two orders. The full picture needs to be reproduced at the surface of a 2D detector, in most cases a CCD. Also, the requirements of resolution and linear intensity (dynamic) range need to be defined.³ The calculation and realization of a 2D Echelle spectrometer is a challenging task that requires the application of most of the parameters described so far. The following example will illustrate this scenario.

Requirement

A CCD of 2048×2048 pixels is among the best suited and will be applied here. The pixel size is $13 \mu\text{m} \times 13 \mu\text{m}$, and the surface area is 26.624^2 mm^2 . The Echelle spectrometer will cover the optimal range of UV-enhanced CCDs, which is at least 220–950 nm. The system resolution will be better than 0.01 nm throughout. The instrument will be designed for laboratory use and must not be subjected to shocks or vibration after installation.

Procedure

To provide good uniformity of light to the large CCD surface, the best possible transfer function is required with minimal introduction of aberrations. Because the pixels are 13 μm , the optimal entrance “slit” will be 26 $\mu\text{m} \times 26 \mu\text{m}$. The most likely slit will be circular with a 25- μm diameter, which will be interchangeable and compatible with the end of an optical fiber. Under optimal conditions, 2 \times 2 pixels will be illuminated by a single optical source. In order to receive the best possible spatial and spectral data, the aberrations will not exceed 50 μm . That, in turn, requires apertures not exceeding $f/6$. The internal angles (ι and ε) will be within 10 deg. It would be preferable to use optical components no larger than 120 mm in the x and y axes to allow availability from several suppliers. As an example, consider $f = 800$ mm and optics with a 120-mm width, which equals $f/6.7$. The Ebert–Fastie setup is known to provide the best performance, so it is used here. A calculation program is then built to process the key equations in parallel because the needed results are the center wavelength, the actual spectral order, the reciprocal dispersion (nm/mm), the interval (nm at 26.6 mm), and the overlap of stripes. After having created the program, the working (blaze) angle Φ and the line frequency of an Echelle grating are entered and the results are reviewed.

Variable parameters include the working angle Φ , grating constant k , Ebert angle ε (beginning with 10 deg), and focal length f . After a few trials, an Echelle grating of 27.245 mm⁻¹ and $\Phi = 46$ deg in an instrument of $f = 700$ mm is found to be suitable. The grating will comprise a ruled area of 120 mm \times 90 mm. The theoretical resolution of the grating will be between $R_p = 0.24 \times 10^6$ in the 54th order at 950 nm and $R_p = 1.172 \times 10^6$ in the 264th order at 195 nm. There is plenty of room, so the grating will not be the limiting element. At 46 deg, the illuminated area will be 83.35 mm, which is calculated with 85 mm \times 85 mm. That will provide an aperture of $f/8.2$, which is well within the picture.

The Ebert mirror must be defined next. The setup needs a light trap in the center of the Ebert mirror, which should be 10-mm wide. The mirror-grating distance will be 0.8 of the focal length, or 560 mm, according to Fastie’s rule. At this point all of the angles can be calculated. The width of the mirror will be 2 \times 85 mm + 10 mm, or 180 mm. The height is 85 mm. The center of both parts of the mirror is 47.5 mm from the center line of the system. The Ebert angles result in 4.85 deg. All numbers are acceptable and promise aberrations $< 50 \mu\text{m}$. If available, a ray-tracing program can be used to confirm the beam behavior. Regardless, the instrument will work at 950 nm in the 55th order and provide a dispersion of 0.655 nm/mm. At 200 nm it will be in the 261th order with a RD of 0.2 nm/mm. The upper limit will be 957 nm. Above that, spectral gaps will appear between orders. Figure 4.31 shows the interval in nm/order and the overlap in nm.

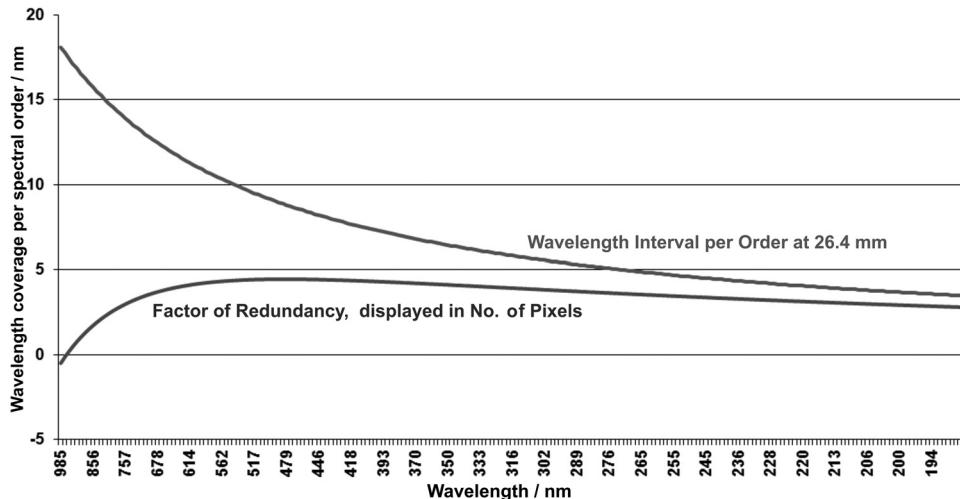


Figure 4.31 Behavior of the spectral intervals at the CCD detector over the wavelength and overlap.

The top curve of Fig. 4.31 shows the interval in nm arriving at the CCD, whereas the bottom curve shows the overlap between two orders. The curves must not touch each other or else the separation would not work. Above the wavelength where the bottom curve (redundancy) crosses zero, data gaps between orders will occur. Both limits define the working range of a 2D spectrometer.

The experimental resolution can be estimated with an entrance aperture of $26 \mu\text{m} \times 26 \mu\text{m}$ by dividing the interval by the number of pixels (2048) and multiplying by 10 (for a safe value). At 950 nm, there is a resolution of $\sim 3 \text{ pm}$, and at 200 nm, it is $\sim 1 \text{ pm}$. Because the intended resolution was 10 pm, the prism can be implemented for the vertical separation of orders. The first step is to look at the construction, applying the parameters already known. The prism is best placed in the collimated beam between the grating and focusing mirror. Prism materials to choose from include quartz (SiO_2), sapphire, calcium fluoride (CaF_2), and magnesium fluoride (MgF_2). Magnesium fluoride is hygroscopic and, like calcium fluoride, much softer and easier to scratch than quartz. Sapphire can be useful because it has higher refractive indices, leading to higher dispersion compared to the other three materials; however, it is not so easy to obtain and is much more expensive. Thus, quartz is selected here.

The next step is to build another interactive calculation table based on quartz data. The variables are the angles A und i_1 . It is important that the median prism output angle is identical to the angle of the arriving light. The vertically dispersed light will deflect up and down from that angle to hold the horizontal plane constant throughout the instrument. Deviations from the horizontal plane will have a large impact on the aberrations. Success is achieved after a few iterations, and a prism is found with angles $A = 20 \text{ deg}$

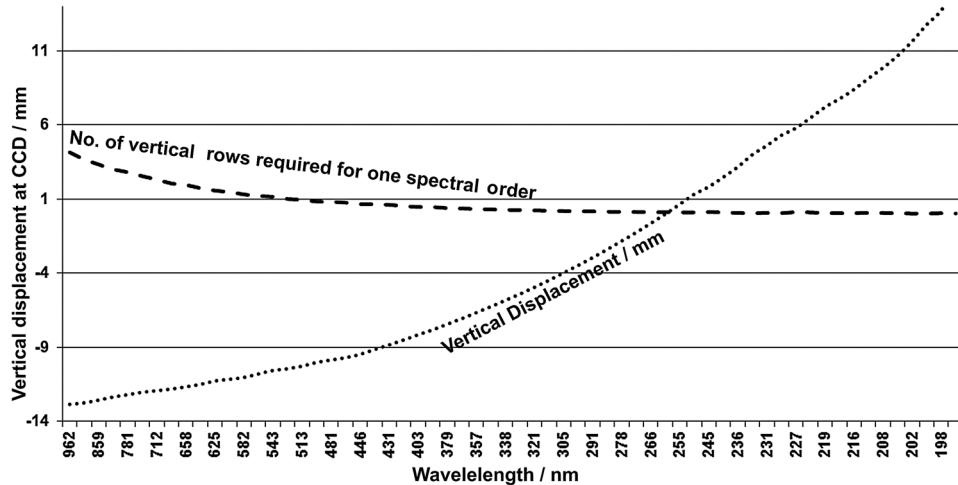


Figure 4.32 Vertical beam deflection by the prism and reserve of spatial resolution.

and $i_1 = 15.15$ deg. Figure 4.32 illustrates the used height of the CCD and if the orders will have a clean separation.

The dotted curve is the wavelength displacement of a 20-deg quartz prism at $f = 700$ mm. The 26.642-mm height of the CCD is almost completely occupied. It is important to note that additional pixel height might be required to cover aberrations at all wavelengths. The presented arrangement ensures separation, as shown in the dashed curve: it correlates the vertical bandwidth per pixel with the height required by one order. At 950 nm, one vertical pixel receives a bandwidth of 4 nm; at 550 nm, one pixel covers 1 nm. At 950 nm, one order (actually the 55th) covers 17 nm, which provides a reserve of factor 4. Consequently, the overlay and aberrations at this point must be kept within $4 \mu\text{m} \times 13 \mu\text{m}$, which is within the target, to keep aberrations below 50 μm . At 200 nm, one pixel “sees” 0.02 nm, and one interval covers 3.7 nm—this provides a reserve of factor 185 to avoid overlay. The critical range is known to be towards longer wavelengths, where the upper curve is above 1.1. It means that one order needs one vertical row of pixels with no additional aberrations. It is logical to make an estimation on the aberrations now.

Equations (2.26)–(2.28) estimate the aberrations and find that $O_1 = 4.33 \mu\text{m}$ and $E = 4.68 \mu\text{m}$, which is good—so long as coma effects from the grating and prism do not add aberrations $> 5 \mu\text{m}$, focal spots of $\sim 30 \mu\text{m}$ are expected in the outskirts of the detector. Furthermore, coma does not contribute to the vertical aberration (it affects the resolution).

4.4.3.1 Concept of a compact 2D Echelle

A basic design can be created based on the parameters found thus far. Considering all of the details, it turns out that it is difficult to keep all internal angles narrow and put the prism in a suitable place without interaction with

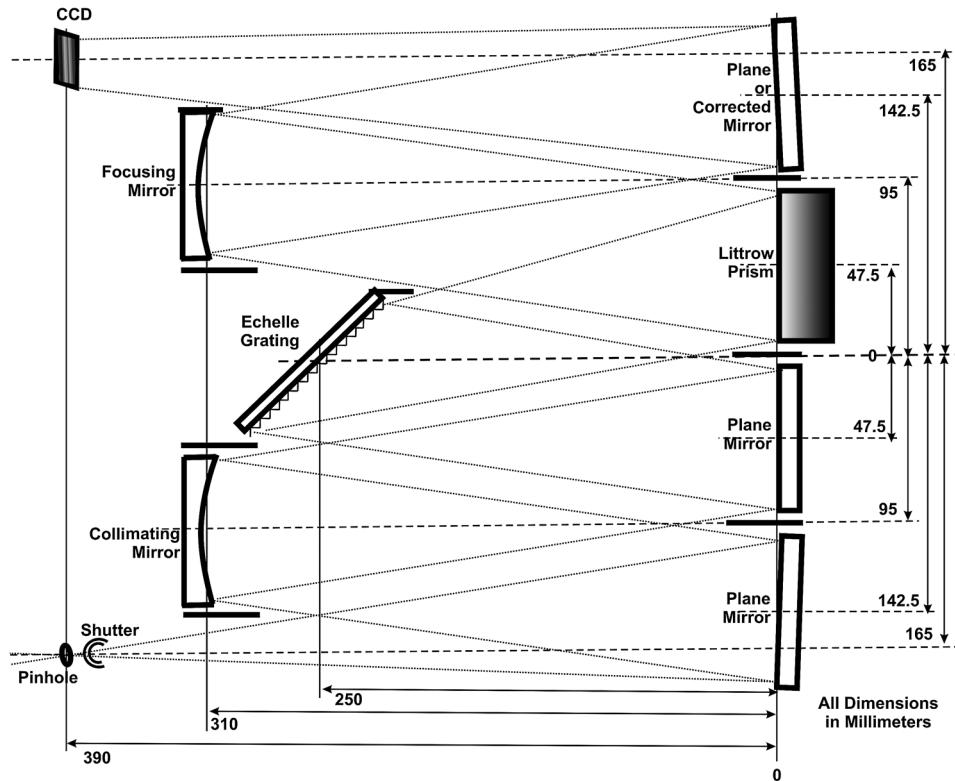


Figure 4.33 A folded, symmetric 2D Echelle spectrometer.

the beam travel. After some modeling, a very compact and stable solution is achieved, as shown in Fig. 4.33.

The system has a compact footprint, occupying less than 45 cm × 45 cm, without the detector. A spatial view of the instrument is presented in Fig. 4.34. If required, a vibration-damped, thermostatic housing can be added. The active mirror surfaces are all 85^2 mm^2 . It is a folded Czerny–Turner instead of the Ebert–Fastie initially supposed. The Fastie rules apply regardless, and all internal angles are less than 10 deg.

The entrance is a pinhole, and an electro-mechanical shutter follows to control the illumination time and avoid data smearing during the CCD read-out. The two curved mirrors are a collimator and a focusing mirror; both are placed aside the Echelle grating. The Littrow prism has its place in the same plane as the beam-turning mirrors. The last turning mirror can be curved to improve the beam quality in the outer regions of the CCD if that scenario is desirable or necessary. Depending on the correction required, a vertical curvature, toroidal, or Schmidt correction can be placed there. Between the components are beam blocking masks and baffles to avoid cross-talk and reduce stray light. The 700-mm focal distance in the entrance arm is split into the following paths: pinhole > plane mirror > collimator > plane mirror >

grating. The grating–curved-mirror distance follows the Fastie rule and is 560-mm long. The exit arm is mirror symmetric, with the prism acting as a vertical disperser and a reflector simultaneously. Both the prism and the optional correcting mirror do not change the focal length of the arm.

4.4.3.2 Comparison of an Ebert–Fastie and a folded Czerny–Turner

In the folded system, the prism–collimator path is 310-mm long, which is the correct distance for the required deflection of all involved wavelengths. Furthermore, both angles ϵ at the grating are slightly wider than first calculated due to the shorter distances. Now $\epsilon = 8.7$ deg instead of 4.8 deg, as planned for the Ebert setup. This difference produces small changes in the grating dispersion.

At first glance, there are two disadvantages: the system is more complicated to manufacture and adjust, and the luminosity is less. However, it carries the advantage that it enables both vertical and coma correction by curving one or more of the plane mirrors slightly, as described for the last mirror. For this setup, all mirrors are considered to have a reflectivity of 0.9, and the prism to be 0.95 in transmission and 0.87 in reflection. The grating efficiency is 0.8. A total T of 0.41 is found at a luminosity $L = 4.75 \times 10^{-5}$. A system without a folding mirror would be 30% more luminous. Because that is within the same order of magnitude, the loss might be acceptable. In comparison to the first calculation (in Section 4.4.3), the prism needs recalculation for the shorter pathway. A Littrow quartz prism with $A/2 = 40$ deg under $i_1 = 31$ deg will provide the required parameters. It will emit a 960-nm wavelength under 28.51 deg, and 195 nm under 33.62 deg, which perfectly envelops the horizontal plane. According to the focal length provided, the CCD will be used between –12.1 mm to +12.8 mm in the vertical. Thus, 24.9 mm of the available 26.6-mm height will be used. The data in Table 4.3 present the performance parameters of the folded, symmetric Czerny–Turner.

The detector receives a total of 213 horizontal spectral intervals (stripes), and each one represents one spectral order emitted by the Echelle grating. The orders below number 55 are not complete. The orders above number 267 can work if an efficient input signal towards the deep UV was provided. Twenty of the 213 stripes are documented in the table. Note how the ratio interval/order drops with decreasing wavelength, resulting in increasing resolution. Each order provides a reserve and can use more vertical pixels than required, as shown in the column *Overlapping pixels per order*. The reserve of pixels (13- μm height) is rather large in the NIR. Towards the green, the dispersion of the grating increases faster than that of the prism, and the reserve drops. Below 550 nm, the prism dispersion changes faster than that of the grating, leading again to increasing reserve. Reserve is required to separate stripes despite the aberrations that occur in the corners of the CCD. Naturally, each stripe has a different wavelength at its beginning and end. Also, both will experience

Table 4.3 Performance parameters of a folded, symmetric Czerny–Turner system (shown in Figs. 4.33 and 4.34).

Wavelength [nm]	Grating order	Covered interval [mm]	Vertical position at CCD [mm]	Overlapping pixels per order	Horizontal tilt angle [deg]	Bandwidth per pixel [pm]
1000	52	18.4	-12.4	—	—	—
950	55	17.7	-12.15	6	-0.5	8.6426
900	58	16.5	-11.93	6	-0.6	8.0566
850	61	15.7	-11.7	5	-0.6	7.6660
800	65	14.7	-11.5	3	-0.6	7.1777
750	70	13.8	-11.33	3	-0.7	6.7383
700	74	12.9	-11.13	3	-0.7	6.2988
650	80	12.0	-10.9	4	-0.8	5.8594
600	87	11.0	-10.52	4	-0.9	5.3711
550	95	10.1	-10.17	4	-0.9	4.9316
500	104	9.2	-9.6	5	-1.0	4.4922
460	113	8.5	-9.1	4	-1.2	4.1504
420	124	7.7	-8.52	6	-1.3	3.7598
380	137	7.0	-7.3	7	-1.5	3.4180
340	153	6.3	-5.85	7	-1.8	3.0762
315	166	5.8	-4.5	7	-2.0	2.8320
290	180	5.3	-3.25	8	-2.3	2.5879
265	197	4.9	-1.18	10	-2.9	2.3926
240	217	4.4	1.7	13	-3.4	2.1484
215	243	4.0	6.7	17	-4.4	1.9531
195	267	3.6	12.81	21	-5.9	1.7578
192	271	3.5	14.3	25	-6.5	1.7236

different dispersion and travel to different heights. Therefore, the stripes are neither horizontal nor parallel to each other. The *Horizontal tilt angle* describes the specific angle of each of the sample stripes at the detector versus the horizon. The rightmost column finally provides the *Bandwidth at one pixel*. The real spectral resolution will be 3–10 times the bandwidth, depending on the wavelength, coma, and position, if no further correction is made. This could be a matter for curved or toroidal folding mirrors, or a Schmidt plate, if correction is required at the end. It would result in the modification of the last plane mirror without changing the rest of the system. A selection of eight orders leaving the prism and arriving at the CCD is shown graphically in Fig. 4.34 beside a spatial view of the system.

The left part of Fig. 4.34 presents a spatial graph of the 2D Echelle system described previously. It is based on a cage of Super Invar® rods. The holders of all optical components are mounted to parallel rails for easy adjustment and solid position. The entrance and exit are more distant to the rear rails; they are part of a massive front plate, which itself is fixed to the four extended side rods of the cage. The entrance aperture, the shutter, and the CCD adapter are accessible and adjustable. The only function of the enclosure is shielding.

The right part of the figure represents the distribution of stripes at the CCD and the principle of the reflecting prism as mounted in the folded 2D Czerny–Turner. It is arranged so that the median dispersed beam in the

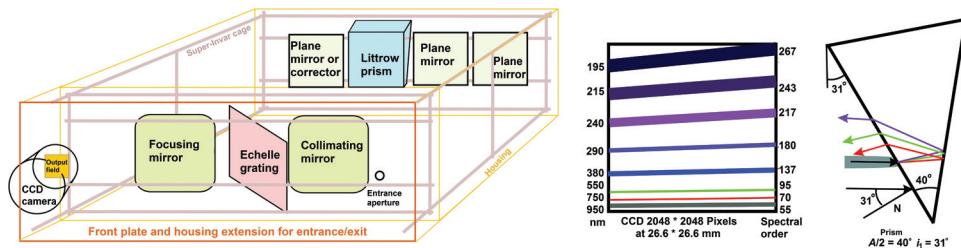


Figure 4.34 Spatial sketch of the 2D Echelle system, and the beam behavior at the prism and the 2D detector.

vertical orientation is at the same plane as the illumination. The prism is tilted so that i_1 travels parallel to the horizon of the instrument and in the center of the grating and CCD. In reality, the prism's front side is fully illuminated with a mixture of all wavelengths, and the dispersed rays also overlay each other to become refocused at the detector. The left part of the figure is the graphical transformation of Table 4.3. The numbers on the left are the center wavelengths of the stripes, and the order number is on the right. Due to dispersion, the lower wavelength on the left is always at a lower angle than the upper wavelength on the right end, while the tilt angle increases towards the UV. For instance, the interval of 17.7 nm in the 55th order has an angle of 0.5 deg, and the 3.6-nm interval in the 267th order has an angle of ~6 deg. The thickness of the stripes is not to scale, but the relation between the different stripes is indicated (the numbers are provided in the table).

The arrival of spectral stripes under different angular orientations and densities has severe disadvantages for the read-out of a CDD detector. All parameters considered, it means that the detector content must be picked in a pixel-by-pixel fashion. (Chapter 5 discusses CCDs and read-out modes in greater detail.) In short, the system control computer needs a software matrix that sorts and recombines pixels and wavelengths. Information stored in a pixel group composes a certain data point. To represent one wavelength in the NIR, it could be a box of 2 pixels \times 2 pixels, whereas in the UV it could be 2 pixels \times 100 pixels. At a read-out time of 10 μ s per pixel (which is required for the best quality of data), the total read time in the example will be 42 s. If photometric precision were sacrificed, a read time of 0.5 μ s would speed the transfer of the total content to 2 s. Other speed/linearity combinations are in between. This is why a shutter is necessary, with the exception of some pulsed applications.

Besides the excellent point-to-point transfer of rays, system stability is important. First, a calculation of aberrations is a good idea. Equations (2.26) and (2.27) indicate that the focus will increase by less than 6 μ m. At f/8.2, coma will be negligible. Because no parts need to move, the final mirror can correct all aberrations. Ray tracing is a good tool to achieve optimal beam travel. If a pinhole of 25 μ m is used, it can be estimated that even without correction, the spots will stay within 30 μ m of the outskirts of the detector, which will be enough to separate all stripes.

Stability is required to ensure that all wavelengths arrive at the same place every time. Drift and other mechanical variations must not exceed the size of half a pixel. Thermal stability is found by Eqs. (2.31) and (2.32). Suppose a chassis made of aluminum and optics based on BK7 glass experience a thermal variation between 20–25 deg, while the system was adjusted and calibrated at 20 deg. The value found by Eq. (2.31) is $dx = 154 \mu\text{m}$. With increasing temperature, the system will “grow” in all directions, which changes the distances and the angles, as well. The defocusing factor found by Eq. (2.32) is $dy = 19.25 \mu\text{m}$. Both values are out of tolerance; even the glass substrates will compensate slightly. At a thickness of 25 mm for all optics, and $K_{Glass} = 0.86 \times 10^{-5}/^\circ\text{C}$, the compensation effect is calculated. Six bodies of relevance are in the system. The grating will remain outside the picture because of its angle, and the prism is treated like the mirrors. The results are $dx_{optics} = -6.45 \mu\text{m}$ and $dy_{optics} = -2.4 \mu\text{m}$, leaving a remaining defocusing effect of at least $12.8 \mu\text{m}/5^\circ\text{C}$, which is already out of tolerance by a factor of 2. There are two options:

- A. Place the instrument in a dual housing, thus providing thermal regulation to within 2°C . That arrangement provides two more advantages—the enclosure housing can be purged with dry air or nitrogen to reduce humidity and corrosion, and the instrument can be fixed in the enclosure by absorbers for vibrations and shocks.
- B. Mount all drift-sensitive components on Super Invar®. A cage made of four rods per side can carry the holders and fixtures of all relevant components. The fixtures and holders themselves would be made of aluminum. Super Invar® has $K = 2 \times 10^{-7}/^\circ\text{C}$ at room temperature, leading to $dx_{Super\ Invar} = 1.4 \mu\text{m}/5^\circ\text{C}$. For the optics, 10-mm-thick glass substrates were chosen. dx_{Glass} becomes $-2.6 \mu\text{m}/5^\circ\text{C}$ (a slight overcompensation), leaving a drift of $-1.2 \mu\text{m}/5^\circ\text{C}$. dy becomes $-0.21 \mu\text{m}/5^\circ\text{C}$. Both values are well within the tolerance limits. After ray tracing and prototyping, some optimization in the fixtures can produce a drift of zero. Prevention against vibration and shock is required anyway.

4.4.3.3 Constructive precautions

It is good to know that there are no parts moving during operation except the shutter. That allows a very rigid, stable, and reproducible construction. The grating and prism should have easy access and be placed on interchangeable, pre-adjusted holders to allow for the optimization of the wavelength range and resolution. The entrance should be a precision coil to allow for the changing of apertures and the installation of optical fibers to the focal plane. The shutter should also be customer-accessible and vibration-damped. The detector needs to be a high-performance camera that is not integrated but bolted on, which requires a standard adjustable flange that can carry $\sim 2 \text{ kg}$ without problems. The CCD chip will need to be cooled to at least -50°C .

(the long read-out time alone requires low temperatures). There must be no feedback from the camera to the spectrometer, neither thermal nor vibrational. Cooling water is not the most elegant way to carry heat away from the CCD, but it might be the most quiet and efficient besides liquid nitrogen. In general, the whole system needs to be massive and free of torsion.

If the decision favors a thermally stabilized dual-chamber system (A), it will be necessary to provide a room for the final adjustment and calibration. Because the internal chamber temperature will be above the ambient temperature, the room should have the proposed-system operating temperature (e.g., 27 °C). The system needs to reach equilibrium before the final work and acceptance. If the system is the folded Czerny–Turner type, it can easily be placed in any orientation, which is of obvious advantage.

If the decision favors a drift-compensated Super Invar® system (B), the cage might be a construction of probably 20 Super Invar® rods that would hold the fixtures of all optical elements, including the entrance and exit. The housing can be a light one—it only needs to be light tight. Eventually, the Super Invar® cage can be spring-mounted within the housing to absorb vibrations.

Measured 2D Echelle examples can be found in this book's companion volume on applications.⁴

4.5 Hyperspectral Imaging

This form of imaging is also called imaging spectroscopy, spectroscopic imaging, or spectral mapping. The different names describe the combination of spectral measurement with imaging capabilities. Over the course of this book, many topics deal with this issue.

4.5.1 Internal references

The spectroscopic interrelations are discussed in Sections 4.1.5 (output fidelity), 4.1.6 (correction methods for spectral imaging), 4.4.3 (2D Echelle spectrometer for the parallel recovery of wide wavelength ranges), and 5.8.8 (read-out techniques such as multi-spectra spectroscopy, binning, and virtual CCD partition). The latter describe the application-oriented use of imaging capabilities, but even the above mentioned names are not used there. More examples can be found in the applications book.⁴

4.5.2 Example of hyperspectral imaging

Besides the example listed in the previous section, the following description might help illustrate the method beyond the already-treated techniques that utilize multiple fibers, one atop another, and 2D detectors. Hyperspectral imaging is often employed with microscopic applications, such as tissue fluorescence, or Raman analysis of solid samples. Small surfaces are typical, and they are not accessible by standard macro techniques. Consider an

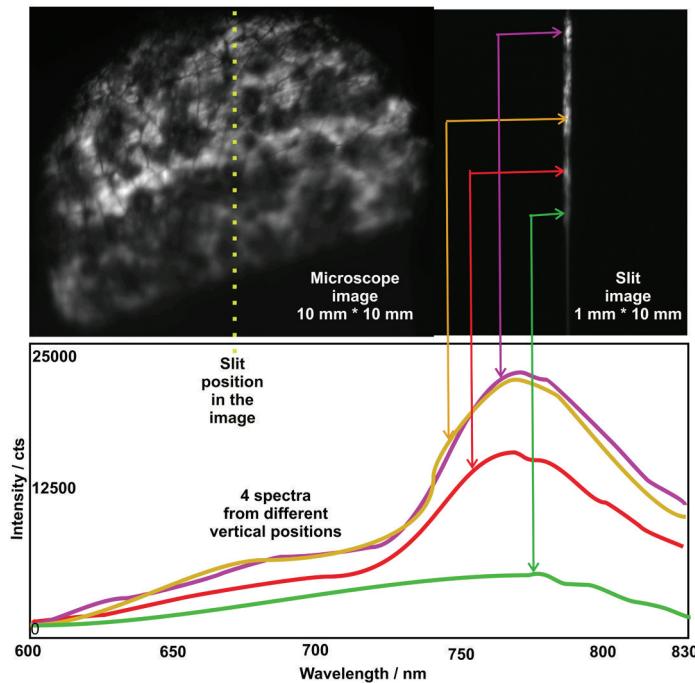


Figure 4.35 A fluorescence image transferred through a spectrograph, slit image, and spectral recovery.

inverted microscope providing a C-mount or similar output for detection systems beside the viewing system. A fluorescent sample was used, and the emission image reproduced by a CCD was $512 \text{ pixels} \times 512 \text{ pixels}$, each of $25\text{-}\mu\text{m}$ size, covering the area of a 12.7-mm square.

The CCD was connected to the standard 2D output of an imaging-corrected, 300-mm spectrograph. The full image is shown in the upper-left of Fig. 4.35; the intensity is presented by the brightness of the signals. The transfer of the full image was enabled by two modifications: One of the three grating positions was equipped with a plane mirror, and the entrance was equipped with two kinds of apertures.

The left image of Fig. 4.36 shows the standard slit, actually opened to 1 mm. The right image displays the $10 \text{ mm} \times 10 \text{ mm}$ wide aperture that enables transfer of the entrance image 1:1 to the detector. The CCD applied was $12.7 \text{ mm} \times 12.7 \text{ mm}$, which is larger than the aperture. Thus, in photo mode the detector is not fully illuminated. Because the spectrograph requires the same illumination angles in the entrance for both the spectral and photo mode, the mirrors and gratings limit the size. Interchanging the slit with the aperture can be done either manually or by an extra stepper motor. The setup offers several further modes of data acquisition: given the slit and a 2D detector, one could have the two detector dimensions be space and spectrum,

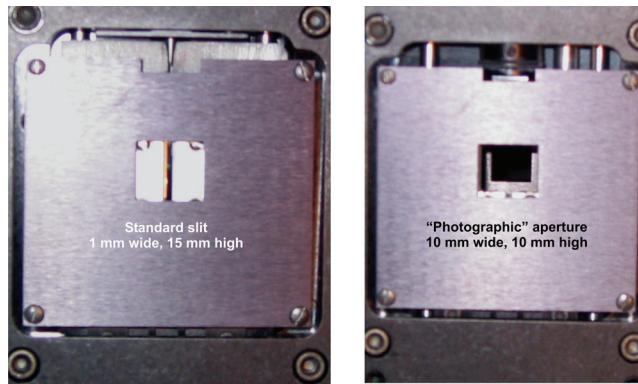


Figure 4.36 Special dual entrance of the “photographic” spectrograph.

while the entrance aperture selects the line in space. One could also have the spectrum versus time for a given point and then raster the scene over the entrance aperture. Or one could have space versus space and use a grating to select a spectral region; this is useful for applications with discrete narrow lines, as in atomic spectroscopy, by providing gaps between the lines.

4.5.2.1 Image reproduction and spectral recovery

After examining the full image of Fig. 4.35, the center (marked by the yellow dotted line) was defined to be the region of interest (ROI) for spectral evaluation. The aperture at the spectrograph entrance was replaced by the slit of 1-mm width. After that, the image looked like that shown in the upper-right of Fig. 4.35. Next, the grating of 150 mm^{-1} was moved in and delivered 230 nm with a bandwidth of 20 nm to the detector. The CCD pixels have been binned by 8 in the vertical direction, producing 64 separated spectra with an interval of 600–830 nm each. Four of those are reproduced in the lower part of the figure.

4.5.2.2 Overlaid hyperspectral image recovery

If the distribution of a narrow wavelength interval distributed over the x - y surface must be recovered, the aperture is left in place, a grating of rather high dispersion is set to a selected wavelength, and a measurement is conducted. Proposing a “slit width” of 10 mm and a 300-mm spectrograph with a grating of 1200 mm^{-1} , the recorded bandwidth will be $\sim 25 \text{ nm}$. In that case, the spectra, which represent different horizontal positions, overlay each other. If each pixel of the CCD is read separately, the vertical information is the true spatial position, but the horizontal information is the convolution of position and spectrum. Even that information is truly “hyperspectral”: the original set of the data in the “cube” will look smeared and need de-convolution, which will not be straightforward but iterative.

4.5.2.3 Separated hyperspectral image recording

In a set-up like that shown in Fig. 4.35, one or many spectra at the same horizontal position can be recovered. The ROI must first be brought to the horizontal center to get its light through the slit. If a handover of camera software and x - y positioner of the sample table is provided, the repositioning can be automated rather easily. Having the right x position in place, the aperture is moved out, and the slit is moved in and set to the desired bandwidth. Because that is also a function of slit width, there is a correlation of bandwidth and recovered width. The vertical size of the ROI is defined by the vertical pixel grouping of the CCD. The full height of the CCD can then be used, and the normal spectral interval and distribution is provided.

Hyperspectral data obtained from surface mapping will lead to 4D data sets. In a typical case, the wavelength distribution will be shown in the x axis, while the spatial position is documented on the vertical and the beveled axis. This arrangement enables presentation of spectra atop each other and leaves space for the display of magnitude. The example shown in Fig. 4.37 contains 64 spectra of different vertical positions.

4.5.2.4 Hyperspectral imaging supported by filters

Beside microscopy, applications such as inspection, surveillance, surface analysis, etc., are based in practice on hyperspectral imaging. Thus, searching for spectral features is not the goal but rather the distribution of probable spectral signals over the sample surface. In these applications, the combination of a 2D spectrograph and bandfilters will save time. In the experiment described here, a spectrum of the spectral range of interest would be taken by either looking at a reduced area or even integrating over the full surface.

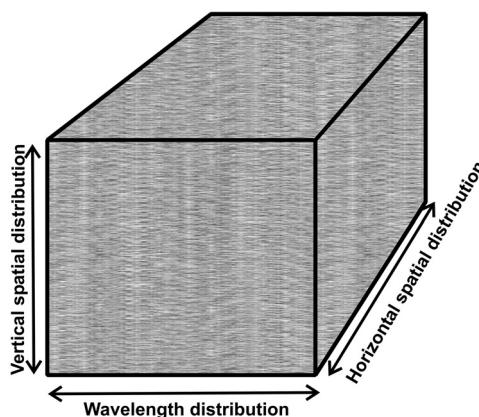


Figure 4.37 The data cube.

4.5.3 General design for hyperspectral imaging

The setup shown in Fig. 4.38 collects light from the object under inspection (e.g., the brick structure), either by a microscope or other optics. The light travels a collimated pathway before it is refocused to fit the spectrograph. Filters can be placed into the collimated part; they can either limit unwanted signals (as Raman notch filters do), let only a certain band pass through, or both. In the case of band-pass filtering, the spectrometer will collect the maximum of the sample area and guide it through the imaging aperture, which represents the focal plane. The object will be reproduced at the collimator, the mirror at the grating turret, and finally at the CCD detector. This setup distributes the wavelength interval, transmitted by the filter, over the surface. After the measurement, the filter can be changed for a different wavelength band. This method is called “photo mode.” If the spectral resolution of spots or vertical slices is the goal, the ROI is moved to the center of the CCD, the slit is moved into position, and a grating is brought to the required spectral position. The detector then receives the spectrum dispersed by the grating and emitted from the ROI. The width of the ROI is defined by the slit width and optional masks, whereas the height is defined by the active CCD area. Whether the height is integrated or resolved is left to the user.

4.5.3.1 Design considerations

The systems described and shown so far all rely on the classical beam travel concept, which requires that parts of the beam are converging, dispersing, and

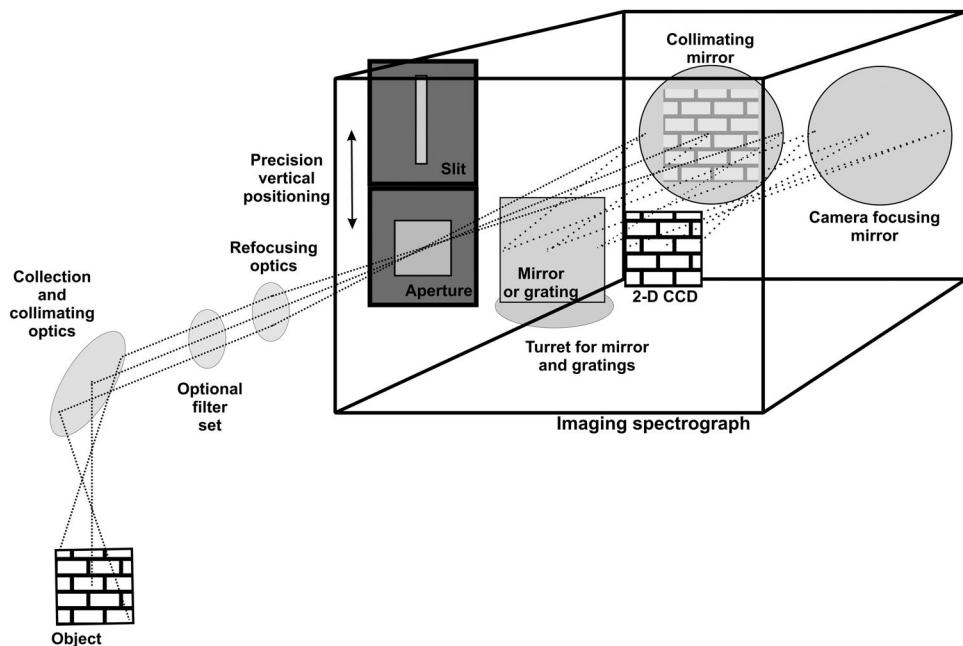


Figure 4.38 Components of a multiple-use spectroscopy and imaging system.

collimated. Because the image reproduction does not happen in a small spot but in a field (e.g., about 10 mm × 10 mm), rather large, high-grade optics are required for good optical quality. Provided that, the image and the spectral components will experience clean transfer and reproduction in the plane of detection. Those requirements are not always guaranteed, and thus some designs work with collimated illumination. Hence, the beam travels in a parallel shape through the spectrometer until it hits the grating, which will create a wavelength-dependent dispersion.

Thus, the first half of the beam travel is parallel, whereas the second half is focusing. To accomplish that scenario, the spectrometer will either be equipped with a plane reflecting mirror between entrance and grating or the beam's cross section will be masked down, typically to less than 15 mm × 15 mm. Both solutions reduce luminosity and fidelity, and make the system a bad tool for standard spectroscopy. So long as the modification is a mask at the collimating mirror, it can easily be removed for “normal” function. Depending on the sharpness of the images processed, the more expensive first version or the second one will be applied.

References

1. M. Lindrum and B. Nickel, “Wavelength calibration of optical multichannel detectors in combination with single- and double-grating monochromators,” *Appl. Spectroscopy* **43**(8), 1427–1431 (1989).
2. W. L. Wolfe, *Introduction to Imaging Spectrometers*, SPIE Press, Bellingham, WA (1997) [doi: 10.1117/3.263530].
3. L. Xu, M. Davenport, M. Turner, T. Sun, and K. Kelly, “Compressive echelle spectroscopy,” *Proc. SPIE* **8165**, 81650E (2011) [doi: 10.1117/12.894191].
4. W. Neumann, *Applications of Dispersive Optical Spectroscopy Systems*, SPIE Press, Bellingham, WA (to be published in 2014).

Chapter 5

Detectors for Optical Spectroscopy

5.1 Introduction

All kinds of detectors are well-documented in general and factory literature, as well as the specifications.¹ It is therefore not necessary to start from scratch but instead concentrate on the specifics important for optical spectroscopy. Surface-area detectors, such as diode arrays or 2D CCDs, are almost always mounted in the focal plane of a spectrograph. Thus, they interact with the other functions of the spectrometer, which means that the detector can create perturbation. Alternatively, single-element detectors, such as phototubes or solid state elements, are almost always mounted behind the exit slit of a monochromator, interacting with the spectrometric functions only in very rare cases. Because any detector has a relevant contribution to the total system performance, the selection of the detector and its specifications should be considered with great care.

5.1.1 Work and power of light signals

The power (P , in watts) or the work (W , in joules) of light is defined by the energy carried by the photons. The value increases linearly with the eV scale. The wavelength scale is inverse and nonlinear to the eV scale. (The interrelations are discussed in Sections 1.2 and 1.3.)

5.1.2 Basic parameters of detectors

The most important characteristics of all detectors are as follows:

- **Quantum efficiency (QE)** designates how many electrons are created per photon (P), per watt (W), or per joule (J). The QE changes for almost all detectors with the wavelength.
- **Gain factor** shows what current or voltage the detector system provides per detected unit.

- **Background (BG)** defines how many electrons the detector produces without receiving input light. The numeric value strongly depends on temperature and time; it generally increases with both parameters. Normally, it also increases towards the NIR–IR range of the spectrum. Cooling will have a positive effect and reduce the background level. Below the level of the background, no reliable data can be expected. Signal levels close to the background level will have rather large uncertainty.
- **Background noise** is the variation of the background signal. If no actuation work (“shot noise,” which is typical for active detectors such as PMTs or MCPs) contributes to the background, the noise (variation) is the square root of the background’s mean value.
- **Linear range** is the signal range, which translates a linear change in photon flux into an also-linear change of the output signal of the detector. The manufacturer data vary in some respect in the tolerable limits of linearity. The most accepted limit stays within 1% of the true value. Often, the linearity curve is S-shaped. In the useful central, linear range, the curve can vary slightly around the pure linear function.
- **Signal noise** comes from the quantum statistics of the arriving photons. The standard deviation (STD) is the square root of the signal’s mean value. Single events such as shot noise can add noise. The maximum deviation (very rare events not taken into account) is statistically seen about five-fold compared to the STD.
- **Acceptance angle Ω** defines the cone angle of the arriving light, creating linear output signals. So long as the beam is not wider than this angle, all light is accepted by the detector. The spectrometer systems discussed here almost always provide narrower cone angles than the detector accepts.

5.1.2.1 Pre-amplifier considerations and wiring

Optical spectroscopy often requires safe data near the very limit of a detector, regardless of type, which in turn requires the best wiring. Wiring between the basic detector and the first stage of the current/voltage converter (or counter, or integrator) should be as short as possible to avoid antenna effects. Leaks between the signal and ground, or supply voltage, can be fatal! Any kind of ground loop will introduce imbalances, interference, and beat frequencies. The first stage of electronics needs to be designed to the current or voltage range, impedance, and frequency limits of the detector element. For instance, a DC-to-MHz pre-amp is not well suited for a photoconductive element optimized for a few hundred hertz because it will add extra noise and drift. In the case of PMTs, the socket is indeed critical.

Three basic kinds of socket wiring are offered: low-noise DC acquisition, high-frequency modulation detection, and photon counting. In the case of array detectors, the element might be able to provide signals overlaid by noise

of only a few electrons. A “general standard” read-out electronic system can add several times the background and noise, destroying the probably excellent performance of the chip. The industry provides optimized sensor electronics, which can appear expensive at first glance but might be worth the price. Detector manufacturers often offer several stages of performance for the detector systems. Finally, free working specialists for low-noise analogue, high-frequency front end, and special electronics offer their services to create optimized front-end devices. In optical spectroscopy, the issues of detector front-end electronics, housing, shielding, antimagnetization, high voltage supply, and gate steering cannot be overrated.

5.1.2.2 General signals and sources of noise in optical detector systems

In Fig. 5.1, the lower green curve is the varying background, which is mainly thermally based. The varying blue curve is the calculated output signal; it is affected by the input light, as well as the background from the detector itself, and the gain. Its fluctuation includes all variations: that of the photons themselves, the background, and eventually the noise coming with the gain. The collected noise in the output signal can exceed 100%. Because output values $>100\%$ are not possible, the sum of the photon signals plus noise drive the system into saturation much earlier than often expected. The averaged value of the integrated data (brown curve) will start to bend rather early, typically above 80%. The red curve represents the progress of the SNR. Even the noise numbers increase with increasing signal, and SNR improves, while

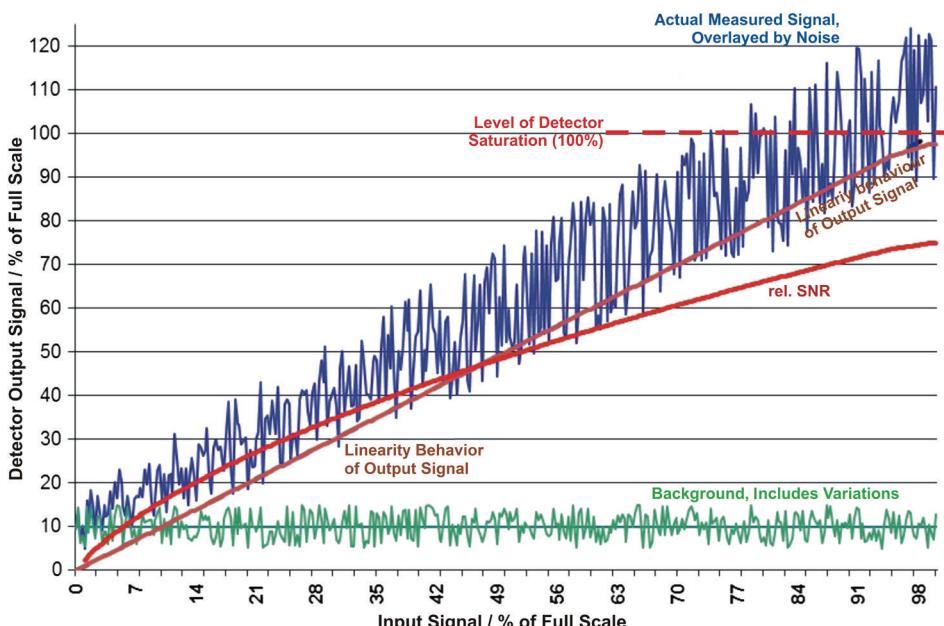


Figure 5.1 General signal parameters of optical detectors.

the influence of background declines. The red and the blue curves also demonstrate that near the background level and near 100%, measurements are unreliable. In almost all cases, the manufacturers leave space in the dynamic linear range of a detector to cover the events described.

5.1.3 Detection limit, noise, and SNR

The general equation for the SNR is

$$SNR = (S - B)/N, \quad (5.1)$$

where S is the average value of the signal, B is the average value of the background, and N is the value of the STD of the noise amplitude. The parameters “detection limit” (DL), “noise” (N), and the “signal-to-noise ratio” (SNR) in an optical spectroscopy system all occur according to two versions: In an absolute measurement, the increasing signal grows from zero (or background) upwards; the increase of the measured parameter is also an increase in the photon intensity. In a relative measurement, the measured parameter relates to a reference of 100% signal. The increase in the measured signal is often a decrease in photon intensity. This difference requires two different views.

5.1.4 Detection limit, noise, and SNR in absolute measurements

In an absolute measurement, the detection limit is defined as equal to the STD of the background. That signal level has an SNR of 1:1. The median-squared variation in a signal equals the arithmetic median value. A detector delivering an average dark signal of 100 units has an STD of 10 units. There are several applications that are limited by the detection limit of the detector system if weak signals are provided; examples include fluorescence and Raman applications. If the experimenter is able to reduce the background of a counting PMT to 4 pulses per second (pps), the STD and the detection limit both drop to 2 pps, which is roughly the best a PMT can perform. If the signal was 100 pps, the SNR becomes 100:2. The STD of the signal becomes the square root of 100, or 10 pps. The STD calculation is based on the Poisson distribution, and this uniform kind of noise is also called “white noise” or “Johnson noise.” The maximum fluctuation of the noise values in both directions from the average value will be approximately a factor of 2.5 of the STD. At an average of 100 pps, single events between 75 and 125 are expected, as shown in Fig. 5.1. With increasing magnitude, the absolute value of the STD will increase, but the SNR only improves with the square root function of the signal. At a level of 10^6 units, the STD drops to 10^3 , or 0.1%. Some types of detectors have to perform internal work to convert the photons into current or voltage: in the case of PMTs, that is the emission work of the secondary electrons; in diodes, it is the interface resistance; and in CCDs, it is the resistance of the well structures. In all cases the extra work creates a small additive noise to the total, called “shot noise.” It makes no difference whether the photons required to create the resulting signal arrive within a long time

interval or by high intensity. Thus, it is obvious that increasing measurement times will lead to the same improvement of the STD and SNR as increasing the signal. Unfortunately, the drift and instability of all components involved will also increase over time, limiting the reasonable maximum time for signal integration. In total, three sources of variations that are not originated by the experiment must be dealt with.

5.1.5 Detection limit, noise, and SNR in relative measurements

The detection limit is given by the STD of the normalized signal magnitude, which is 100%. Therefore, it is important to note what numbers are given in voltage (or current, or counts) for the reference value and how much time it took. The square root of that value defines the best possible SNR. The variation in a signal (here, the reference minus the background) correlates with the square root of the arithmetic average value. If the reference is 1000 units, with a background of 100 units, the reference itself already carries a STD of $30 + 10$ units, leaving little room for good results. The detection limit is influenced by many parameters; the main sources are the light source (lamp, laser), sample (roughness, turbulence), beam guidance, spectrometer (stray light, losses), and detector (background, sensitivity, dynamic range). The detection limit always plays a key role. The best practical SNR can probably be expected like so: If a background SNR of 10^{-6} is presented for a signal of 1% of saturation, that means, in turn, that the background noise is only 10^{-8} of the saturation signal. Thus, the STD at 100% signal will only be 10^{-4} , and at a 1% signal, the SDT will still be only 10^{-2} of the measurement result. It is out of the question that, at a constant signal amplitude, increasing the experiment times in the principal improves both the STD and SNR. However, compared to absolute measurements, the time limits are even narrower because more components contribute to the result (indeed, at least two measurements need to be acquired for one final result).

5.2 Single-Point Detectors

5.2.1 Phototubes

Also called photomultiplier tubes (PMTs) and secondary electron multipliers (SEMs), phototubes use the effects of absorption combined with the creation of secondary electrons. Arriving photons reach the photocathode, made of special material; if it absorbs a photon, one or more secondary electrons leave the cathode. Inside the tube are several grids called dynodes. In most cases, there are 8–11 dynodes arranged in an electrical series, with increasing voltage towards the anode. The dynodes multiply the secondary electrons several times before the current reaches the anode. From there the current is lead to a measurement circuitry. Even the PMT is among the oldest optical single-point detectors: it belongs to the most flexible class of devices, providing superb detection limits. The quantum efficiency (QE) of the available cathodes reaches up to 0.25 in the wavelength range of ~110–900 nm. The lower

wavelength limit is given by the window of the tube, and the upper one by the absorption edge of the cathode. The lower the photon energy is, the lower the probability of absorption.² That dynamic produces fewer secondary electrons that can be multiplied and analyzed, and the remainder create heat. A Japanese manufacturer provides special PMTs for $\sim 700\text{--}1700$ nm and even longer, but at a QE of <0.01 . For the UV and visible range, one can calculate with an average QE of 0.1 if no data sheet is available.

The QE will either be defined in photons or A/W (over wavelength). Whether the notification is in amperes per watt or the inverse, it might be important to know the power W of a photon at the wavelength considered. The conversion is shown in Fig. 1.2. The current A emitted by the cathode is provided without any multiplication factor. Normally, the data sheets also contain the dark current (background) as a function of temperature, and the saturation current of the cathode. With the help of those data, the dynamic range can be calculated.³

The great variety of PMT types provides a wide range of parameters (see Figs. 5.2 and 5.3). For example, the maximum high voltage (HV, x axis) can be <700 or >2000 V. The output signal (black curve), as a function of HV, follows various shapes. The background (dotted curve) and the noise (dashed-dotted curve) grow exponentially with the high voltage. The SNR (dashed curve) first increases with the high voltage and the signal amplitude, but at $\sim 2/3$ of the maximum HV, the SNR curve tilts because the noise starts to grow faster than the signal. The best linearity can even be below the best SNR position, perhaps between 0.3–0.5 of the maximum HV. PMT detectors are sensitive to input strain; a concentrated light beam can, even at switched-off HV, hurt the

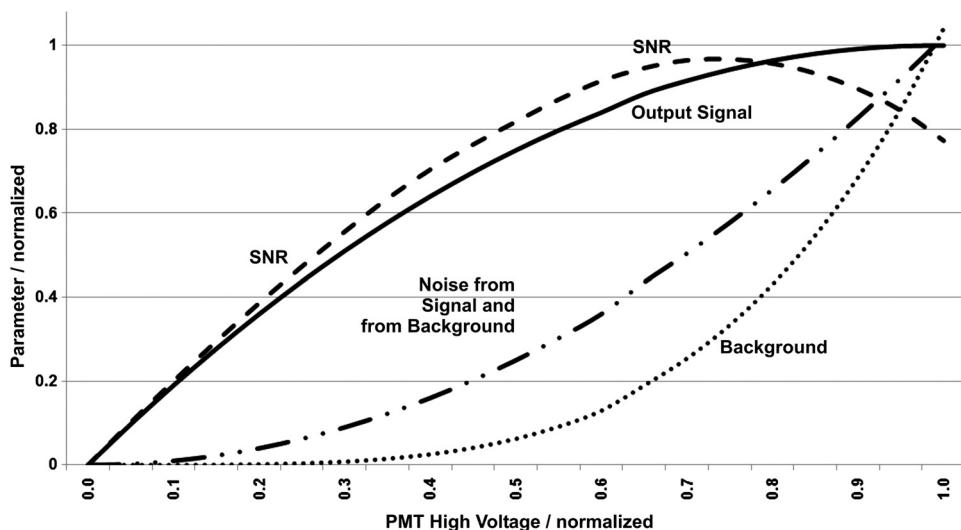


Figure 5.2 General and normalized data curves of phototubes (also valid for other detectors with variable internal gain).

cathode. Strong cathode overload can lead to losses in linearity, increasing the noise. The larger the area of the photocathode is, and the deeper it can “look” into the red part of the spectrum, the higher the background and therefore the noise. Cooling will help and drastically reduce the background if the upper wavelength limit is above 600 nm. On the other hand, the temperature must not be too low because of two reasons: the thermal strain in the system can lead to damage, and low temperatures will increase the emission work of the cathode. Apart from some special versions, the best cooling is reached between -15°C and -25°C . The smaller the cathode is, the smaller the background signal. Thus, PMTs for photon counting often have a much smaller cathode than standard products. The illumination (see Section 5.3 and Fig. 5.4) should be arranged in such a way that as much of the cathode area is used as possible. This will minimize the danger of local overload and compensate for sensitivity variation over the surface.

PMTs are unipolar systems. The output signal cannot undercut the zero value. For further data processing and for background reduction, it is important to cover all variations. If the background is perfectly tuned to zero, the smallest drift or noise towards the negative will destroy the statistics and create false data.

Typical PMT efficiencies are presented in Fig. 5.3. The axis of efficiency can show the QE and reach up to 30%. Often, the equal specification A/W is found, showing the current per light power created in the tube. Hence, the factor of photon power W is included in the efficiency, changing the curve shape in a way that shifts the maximum towards the UV. The displayed curves are only for general information.

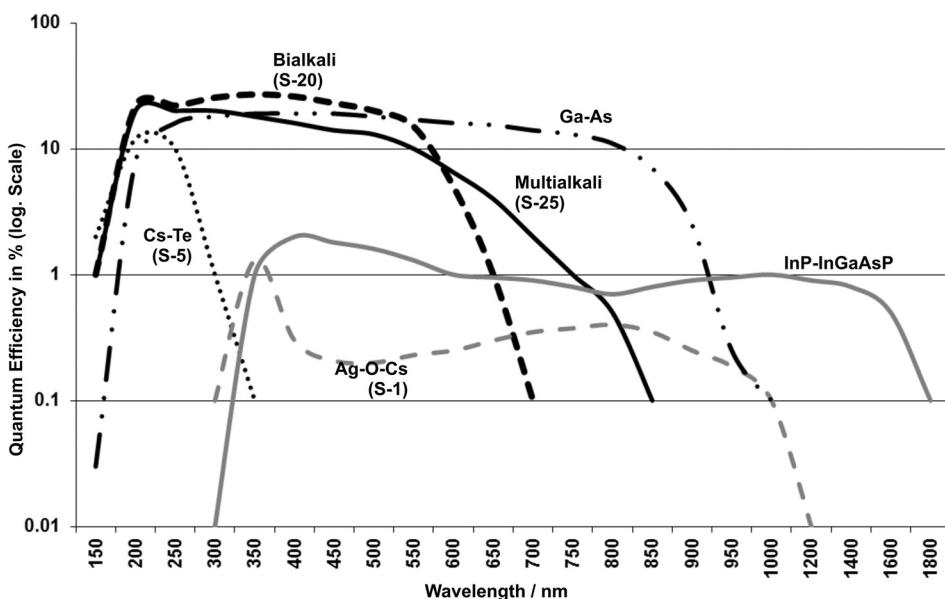


Figure 5.3 Typical quantum efficiencies of common cathode materials.

5.2.2 Comments on the interpretation of PMT data sheets

- The saturation current of the anode is typically in the range 0.01–0.1 mA.
- The typical sensitivity of the cathode is 1–100 mA/W.
- The maximum current amplification is in the range of 10^2 – 10^7 , and it is influenced by the number of dynodes in the tube and the amplification voltage (HV).
- The dark current of the cathodes, at ambient temperature, reaches from 1–1000 nA.
- Depending on the type and size of the cathode, the dark current will drop as much as three orders of magnitude by appropriate cooling.
- The maximum allowed dynode voltage between the anode and cathode is 700–2000 V; the useful range in most cases starts at 100–150 V.
- The background signal is also called the “noise equivalent power” (NEP) and is the quotient of the output signal/gain [typically in the femtowatt (fW) range].

5.2.3 A sample calculation for PMTs, valid for an integration time of 1 s

The upper end of the linear range of the anode output curve is proposed to be 50 μ A, and the current amplification at 1000 V HV will be 10^6 . A photo current of 50 pA (50μ A/ 10^6) is required to create the maximum output current. If the conversion sensitivity of the PMT is 50 mA/W, 1 nW of photons needs to be applied to get an anode current of 50 μ A. If the dark current is 50 nA, the SNR will be 1:1 at a 1-s integration time. That current, in turn, equals a photon power of 50 fW. Assuming a wavelength of 400 nm, one photon carries the power of 5×10^{-19} W; thus, a light stream of 10^8 photons/s is required. For comparison, a PMT equipped with a multi-alkali cathode and cooled to -25 °C is assumed, which leads to a SNR yield of factor 100. This means that the detection limit at a 1-s integration time drops from 50 pA to 500 fA, or 10^5 photons/s. In the (realistic) example, the uncooled PMT will have a linear dynamic range of 100μ A/50 nA, or 2×10^6 . By cooling, the linear range is increased to $>10^8$. In both cases, the variation of the high voltage is not yet included. Varying the HV is reproducible and can lead to a further increase of the dynamic range by 10^2 – 10^4 . Applying all variables, a PMT can work over a linear range of up to 10^{12} . That wide working range is not performed by any other type of detector in optical spectroscopy. The PMT is the most flexible single-point detector and is available in hundreds of versions (many developed for specific applications).

5.2.4 Photon counter

Each single photon, if detected by absorption, creates an electron that will, after the gain stages, arrive at the anode in the shape of a cloud (also called pulse).

This behavior permits counting the events via a method called “photon counting.” Special PMT types that incorporate dedicated dynode structures provide sharp pulses at the anode, which implies operation at a carefully selected, best-adapted dynode HV, reducing the total dynamic range. Because the following discrimination electronic circuitry needs a certain amount of time to recover after each pulse (called the recovery time or dead time), typically 50 ns, the counting rate is limited. If a pulse arrives during the recovery time, it is lost in most cases; however, electronic systems are available that count twice in that case.

Over time, photons arrive in a statistical distribution. Therefore, 50 ns is not equivalent to 2×10^7 pps, but distinctively less. A maximum of 10^6 pps is a good assumption. Expressing this value in hertz is definitively wrong because the photons will not arrive wave-shaped. The detection limit is given by the background signal, which is thermally produced. Some of the counting tubes are available with rather small cathode surfaces to keep the background low. Very good counting PMTs lie in the range of 40–1000 pps [also called counts per second (cps)] at room temperature. With the help of cooling, the background quickly reduces. At $-25\text{ }^\circ\text{C}$, a value of 3–100 pps remains. Because the background defines the SNR, and thus the detection limit, a low background is very important. Depending on the cathode material, cooling can also mean a reduction of the QE (valid curves are available from the manufacturers). A good compromise in most cases is found around $-20\text{ }^\circ\text{C}$. The further the cathode is sensible towards the NIR, the higher the background level. Thus, a PMT should not just be selected broadband, but rather it should fit the application. So-called “solar blind” tubes, which work only below 360 nm, need no cooling at all. The dead time can be shortened to ~ 10 ns by carefully tuning the socket wiring, the cabling to the amplifier/discriminator, and the electronics therein, which increases the maximum count rate accordingly.

5.2.5 UV PMTs and scintillators

PMTs are generally closed and evacuated systems. In most cases, the housing is a glass bulb, so the window is one of the limiting parameters. Towards the UV, the transmission of glass reaches ~ 320 nm, standard quartz ~ 180 nm, special quartz 160 nm, and MgF_2 or CaF_2 windows reach 115 and 106 nm, respectively. Open PMTs, with a CsTe or CsI cathode, are also available and work down to 30 nm. They need to be kept under vacuum, and they are sensitive against arriving ions. An alternative for the range < 200 nm involves scintillators, which absorb VUV photons and convert the wavelength by emitting in the near-UV. They can be obtained as a gel (such as barium fluoride) or foil (such as potassium silicate). The efficiency of those materials varies between 30–300 nm (4–40 eV) at values between 10–90% QE. They emit between 300–450 nm; some can even be detected by solar-blind PMTs.

5.3 Illumination of Detectors, Combined with Image Conversion

One of the advantages of a PMT is the large size of the cathode, which allows for high luminosity and large slit apertures. On the other hand, it might happen that the available surface is not fully used by the spectrometer, in which case the following disadvantages apply: (a) every detector surface is inhomogeneous in certain limits (using the complete area helps homogenize), and (b) photocathodes are sensitive to high photon densities. If signals arrive concentrated at the same place over a long time, the cathode can suffer locally, which leads to losses in sensitivity and linearity at that particular spot. In addition, the background emitted from there might increase. Beam widening, optionally combined with a modification of the image, is an efficient way to avoid those impacts. As discussed later in this chapter, it also happens that an image needs to be minimized to avoid overillumination of the detector, which also needs to be avoided.

Some methods of coupling are shown in Fig. 5.4. The dark, shaded areas in the front represent the output slit of a monochromator. From there, the light travels to the right-rear, where the active detector surface is shown grey. The four solutions are as follows:

1. **Direct coupling:** Front-left is the slit, back-right is the receiver, and the circle represents the active detector surface. The output beam is fan-shaped, and the angles are defined by the spectrometers. If the detector has the right distance, at least one dimension will be fully illuminated.
2. **Lens coupling:** Reduction of the image's slit height while not affecting the slit width. The first lens collimates the beam, and the second one refocuses to the detector. The cylindrical lens compresses only the beam height. There will be no sharp image at the detector.
3. **Mirror coupling:** Two reflecting mirrors, both mounted under 45 deg, reimage the slit with modification of the height. The first mirror is spherical and collimates the light. The second one is cylindrical and condenses the height only. A sharp image can be adjusted. Single-mirror versions are also available, based on an elliptical mirror, requiring a 90-deg configuration.

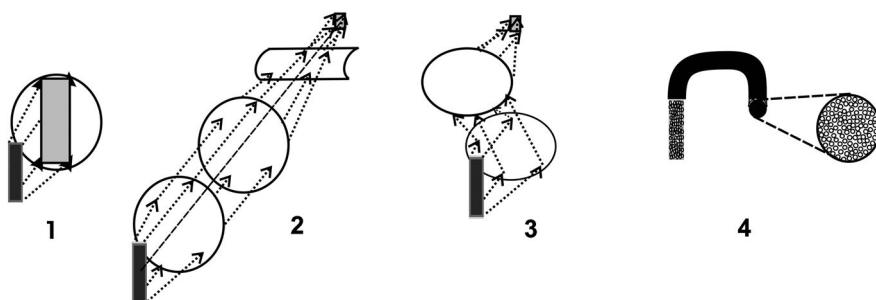


Figure 5.4 Four (out of many) options to couple detectors and spectrometers.

4. Fiber optic coupling: Fiber light guide cable with image converter. The left side of the fiber cable is illuminated by the slit. Optionally, the cable itself can replace the slit and work like a fixed slit; it can be placed very close to the slit; or the optical transfer can re-image the slit at the cable surface. The cable output (right side) can be shaped to the detector surface. In certain limits, the total surface area can also be modified.

Because the detector normally has no impact on the spectrometer function, it can be mounted under a tilt angle in order to increase the used surface. An undesirable exception can occur and create interaction if, for example, the slit surface is very large, and the detector can back-reflect into the spectrometer. Tilting will also help avoid that scenario. PMT cathodes are most often larger than the monochromator slit. A $10\text{ mm} \times 30\text{ mm}$ cathode area is often found. Mounting the detector at an appropriate distance will adapt the surfaces, as shown in image 1 of Fig. 5.4. Cooling systems often require a rather large distance between the detector surface and slit. This requirement can lead to over-illumination, especially when a photon counter is used; it will require transfer optics, as shown in images 2 and 3.

5.4 Channeltron® and Microchannel Plate

A different type of electronic gain is utilized by the Channeltron® and the channel plate. The Channeltron® is a small, snail-shaped tube (left image in Fig. 5.5) that has a coating on the inner surface. The coating releases secondary electrons (e^-) upon impinging electrons and quanta. Along the channel, a high voltage is applied, which creates more collisions and stronger acceleration. After several collisions, a multiplied pulse of electrons leaves the channel. Channeltrons® operate in the energy range of electrons and photons of high energy ($>5\text{ eV}$). They are open systems and need to be placed in vacuum. In most cases, the detector at the end of the Channeltron® is a positively charged metal plate. Upon the arrival of a cloud of electrons, the charge experiences a sudden collapse that can be detected and counted. The entrance of the Channeltron® is grounded to avoid deflection of the quanta. Applications of Channeltrons® are in vacuum and UV spectroscopy, covering energies beyond the range of PMT cathodes that would be either transparent or damaged if used.

As shown in Fig. 5.5, the Channeltron® is only a few millimeters large. The substrate consists of an insulating ceramic tube. The inner surface is coated with a material to create secondary electrons (SEs). The high voltage U_M between the entrance and exit, where the detector D is placed, accelerates the arriving electrons or photons, and multiplies them. The clouds of electrons change the charge of the detector, and the resulting pulses are counted.

A large selection of microchannel plates (MCPs) are available, as well as good documentation and literature. The term MCP can mean two things: (a) the plate itself, or (b) a complete image intensifier system with the plate included.

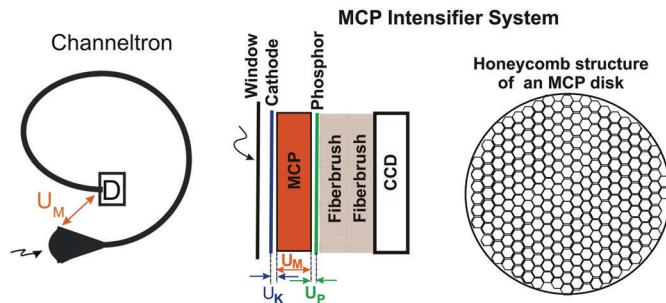


Figure 5.5 Channeltron® and microchannel plate intensifier.

The right image in Fig. 5.5 illustrates these devices. The basis of a MCP system is the plate itself. It consists of a glass disk that is less than 1-mm thick. Via etch lithography, very small and dense channels run through the glass. In most cases, the channels are of honeycomb shape, having a diameter of 4–25 μm . The walls between the channels are only about 1–2- μm thick, so the surface efficiency exceeds 80%. The channels are coated by a material that creates SEs. An acceleration voltage U_M (typically 300–1500 V) generates the secondary effect, which defines the gain factor of the SE process. An arriving and detected electron, or pulse, will collide with the wall several times and leave the channel as a cloud. The gain factor is between 3 and 10,000. The basic MCP only accepts electrons and high-energy photons. The linear range is limited in a dual sense. One limit is the capacity of a channel: the wider the cross-section is, the higher the capacity. The other limit is the total capacity of the plate: because it is charged like a capacitor, breakdown can occur at high currents. Near the saturation current, the signal can become nonlinear or experience a total breakdown. Thus, the MCP can stop working for hundreds of milliseconds; this can happen if a rather large signal is distributed over a large part of the surface. “Overload” phenomena like that are reversible and leave no long-term traces. Harm can be created by electrons of high energy and high speed (as in mass spectroscopy), but not in the range of applications discussed here.

In contrast, energy carriers of <4 eV (>310 nm) are not able to create secondary electrons. Therefore, the entrance of a MCP for optical spectroscopy is a photocathode. The parameters are like that of PMTs, making the QE function of the MCP system PMT-like. The gain and noise functions are also comparable to those of Fig. 5.2. In Fig. 5.5, the cathode is marked blue. To ensure that the e^- emitted by the cathode will precisely enter the opponent channel, a target voltage U_K is applied. The cathode material is carried by a thin metal sheet or quartz substrate, which is ~ 100 μm apart from the MCP. At about -200 V at the cathode, the e^- will pass the substrate and hit the MCP exactly in the opposite place. This method is called “proximity focusing.” If the voltage U_K at the cathode becomes positive (ca. + 5 V), the electrons will drain off. The switching of that optoelectronic port is called “gating”: time slices

reach down to 2 ns or even faster. Modulation of U_K is also possible, providing many methods of nonstatic measurement techniques. In both cases, because nothing is perfect, there are limits. The control voltage U_K is applied in a ring shape around the photocathode. The larger the MCP diameter is, the longer it takes to distribute and equilibrate the voltage step over the cathode. If the cathode rides on a metal sheet, the progress is faster compared to the quartz substrate version.

In any case, the outskirts will respond faster in both directions toward opening and closing the gate. At fast gate speeds, the delay can become critical. In terms of efficiency, the absorption of e^- in a metal substrate (mostly nickel) is $\sim 40\%$ stronger compared to the quartz carrier, which is at the expense of the total performance. There is a trade-off between speed and efficiency (= SNR). In optical spectroscopy, MCPs with diameters of 6, 12, 18, 25, and 32 mm are used. Clearly, a 12-mm MCP is faster than a 25-mm version, and the nickel version is $\sim 10 \times$ faster than the quartz model of the same diameter. The repetition rate of the gate function is also limited due to the rather high switching currents. Depending on the manufacturer, model, and experiment time, the maximum repetition rate is 1–500 kHz. Photocathodes are semi-transparent layers. Photons of >4 eV (<310 nm) will not all be absorbed in the cathode. Some might reach a MCP channel; they create secondary electrons, even with the cathode turned off. Those “gate-off” signals add to the desired “gate-on” signals and increase the measured values. To avoid that scenario, systems are available that also switch the MCP voltage U_M in parallel to the cathode. That technique is called “tandem” or “bracket” gating. For electrical reasons, the gating of the MCP itself is much slower than that of the cathode, in the range of 100 ns. This design makes the whole thing complicated because the MCP must already be open when the cathode switches on, and both will close at about the same time. Because the MCP time function cramps the cathode function, the synonym “bracket gating” is common. It is easy to realize that triggering such a system is anything else than primitive. Both high voltages that steer the cathode and that control the MCP can also be modulated. If, for instance, the MCP gain is sine modulated and the cathode is gated, the system is able to “freeze” a certain point of the sine frequency for read-out and analysis, which permits multi-dimensional frequency response analysis (FRA) of an excited sample or the system itself. The parameters are frequency/time, amplitude, phase, and wavelength. (The luminescence chapter in this book’s companion volume presents further details about the technique.⁴⁾

To the outside world, the cathode either sits behind a window or a fiber brush transports the photons and composes the vacuum feed-through. A brush of fibers is also referred to in the literature as a “bundle,” “faceplate,” or “taper.” Because those expressions are used in other ways, “brush” is used here to mean a one-to-one connection. One side of the fiber bolt is directly integrated into the cathode, while the exposed side of the bolt can either be

illuminated or another fiber brush can connect and transfer the photons almost loss-free. Also, an additional scintillator can be applied there.

The electrons leaving the MCP can be detected by several different means. They can hit a charged metallic plate, composing a position-sensitive detector (described in Section 5.9.2). Also available are combinations of MCP and PMT, where the MCP replaces the cathode of the PMT. This system's advantage is the gate capability. However, the designs most often found in optical spectroscopy are the following combinations. The electronic signals amplified in the MCP are converted into visible light again. To do that, the e^- are proximity focused from the MCP channels onto a phosphor (marked in green in Fig. 5.5) that emits an intensity-linear signal of a color between blue and yellow. A blue-emitting phosphor decays faster but is less efficient; a green or yellow version decays slower but is more efficient. In order to keep the electron beam sharp and to hit the phosphor with high energy, the electrons are accelerated by several thousand volts (U_P). The phosphor emits through a window or, like the entrance does, through a fiber brush. The final detector can be selected from several devices, but in the majority of cases, it is a CCD equipped with a fiber brush entrance. Even the phosphor itself can be selected from a variety of choices—its emission does not hit the CCD sensitivity curve close to the maximum. The more towards yellow the emission of the phosphor turns, the more photons/electrons it creates and the better it fits the CCD response, but the longer the decay time takes.

Two more phosphor parameters need consideration. The emission works in a ball shape: from the point of excitation, the photons travel in all directions, including those outside the detector's view. Neighborhood phosphor molecules will be excited, even at reduced efficiency, and start to glow. That crosstalk increases the output light spot. Furthermore, the decay time of all phosphors increases with decreasing temperature, which is contrary to the advantages of cooling the cathode and the MCP in order to reduce the thermal background of both. Often, the phosphor decay time is defined so that the time between the end of excitation until the decay is down to 1% of the starting value. A yellow phosphor at $+20\text{ }^\circ\text{C}$ will decay in the range of 5 ms. The time typically increases by a factor of 10 per $-10\text{ }^\circ\text{C}$. At $-20\text{ }^\circ\text{C}$, the decay time is perhaps 50 s. For almost all experiments, that time is too long, which is why MCP systems with phosphor output are almost only moderately cooled. A version of MCP read-out, rarely found in civilian applications, is the direct coupling of a CCD after an MCP. The CCD efficiency for electrons is very high—no cross-talk will occur, and no afterglow, either; the whole system performance is perfect. The only problem is that the accelerated electrons hurt the CCD pixels badly. After a few millions of e^- , the efficiency drops vastly, and the linearity also worsens. It is possible to restore the CCD to $\sim 70\%$ of the original performance, but after perhaps five “cycles” the CCD will not be useable. To make matters worse, parts of the MCP–CCD package must

always be kept under vacuum, and special tools and environment are required to revitalize the system. For all that, the direct coupled systems are very seldom found in optical laboratories.

Systems with a phosphor output but without a fiber brush provide the option to collect the emitted light with a lens system and transfer it to a remote camera. Placing two objectives with different focal lengths face to face allows for the spread or condensation of the image without losing sharpness. This method has two more advantages: the CCD camera can be placed in a possibly better surrounding while the MCP is involved in the experiment, and the camera sensor can be cooled to the optimum without affecting the phosphor. At photon energies >5 eV, secondary electrons are directly created in the MCP channel. For high-energy experiments, this fact provides the potential to use an MCP without a cathode or window; the system could be placed into vacuum directly.

5.5 Intensified PMT and Single-Photon Counting

A common method that uses counting and pulsed systems is that of time correlation, also called “single-photon counting” or “time-correlated single-photon counting” (TCSP). It is used to analyze lifetimes in the ps–ns range. A short, exciting light pulse (typically from a pulsed laser, a xenon flashlamp, or a similar stroboscopic, synchronized source) is applied to the sample. Parallel to the flash, a photon counter is started. The arrival of the first detected photons (the first pulse from the PMT) will stop the counter, and the elapsed time will be written into a register. Because the photons and the correlated pulses arrive statistically distributed in time at the discriminator, the rise time and/or decay time of the experiment will appear after many repetitions. The MCP–PMT combination extends those kind of applications by either setting a time window for the active PMT, limiting the analyzed time span, or by modulating the detection. The final signal will then contain the parameter’s frequency and phase for further calculations. The technique is used in fluorescence and similar applications.⁴

5.6 Solid State Detectors

Single-point solid state detectors—such as silicon, germanium, indium-gallium-arsenide, and others—use the effect of charge creation in the crystalline structure upon photonic absorption. Arriving photons can either load or discharge the interface layer of a diode. The process is linear with respect to the photon power and the number of photons. The efficiency depends on the wavelength, the material constants, structure, and doping (endowment). If the absorption of light has a charging (creation of free electrons) effect, the current can be monitored, and the detector will deliver a signal analogue to the intensity of light basically independent from the frequency behavior of the input signal. Detectors of that family are called

“photovoltaic” (PV). A second family of elements changes its resistance upon impinging photons. Typically, the resistance is reduced with the photonic power arriving. Detectors like that are called “photoconductive” (PhC). The effect can be a linear or an exponential one. Either way, the detector element needs an external source to provide the desired signal. If a constant voltage (DC) is applied, the current through the element will change with the intensity of light. Because the electronic sensor system would be electrically parallel, separation is required. There are several methods to achieve that separation, but because the signal changes or minor changes in optical power will be very small, the separation in optical spectroscopy is almost always realized by modulation (AC) and capacitive coupling of the sensing electronics. The driving voltage (DC) is called the “reverse voltage.” To isolate the two voltages, the measurement light needs frequency modulation, which makes stationary measurements impossible but also suppresses all kinds of DC or low-frequency signals, including drift. Some solid state detectors are available in PV and PhC versions.

- **Silicon** (Si, PV) absorbs between \sim 320–1150 nm. The upper wavelength limit is called the “bandgap.” It is not a sharp border, it looks more like a filter curve, and it depends on the previously mentioned parameters and the temperature. Towards longer wavelengths (lower energies), Si becomes transparent. Through variation of the materials, the shape and lower limit of the efficiency curve can be varied to below 200 nm.
- **Indium-Gallium-Arsenide** (InGaAs, or In-Ga-As, abbreviated IGA, PV) offers a vast variety of versions. It is offered for 700–1700 nm, or up to 2200 nm; even versions above 2600 nm are available.
- **Germanium** (Ge, PV) has a useful range of \sim 700–1800 nm.
- **Indium-Arsenide** (In-As, PV) can be used for 800–3500 nm.
- **Indium-Antimonide** (In-Sb, PV & PhC) is efficient between \sim 1000–5500 nm.
- **Mercury-Cadmium-Telluride** (HgCdTe, or MCT, PhC) is also available in many versions that work from 700 nm to 3 μm , or up to 5, 6, 8, 12, or 16 μm .
- **Pyroelectric** (PhC) detectors feature a purely thermal behavior; they provide an absolutely uniform function over the spectrum.

Figure 5.6 shows some of the solid state devices. There are many more, such as lead-sulfide (PbS, PhC), lead-selenide (PbSe, PV, and PhC), etc. In practice, all 3–5 compounds are candidates for NIR–IR detection.

The abscissa is shown in A/W . To imagine how the curves would look if efficiency were presented in photons, the grey dash-dotted line helps illustrate the relative power W per photon. To convert a curve to A/photon , it needs to be multiplied by the power of a photon (rel. P, grey-dashed) and a constant factor. The conversion would obviously shift the maxima of all efficiency curves towards lower wavelengths. The wider the spectral efficiency range of a

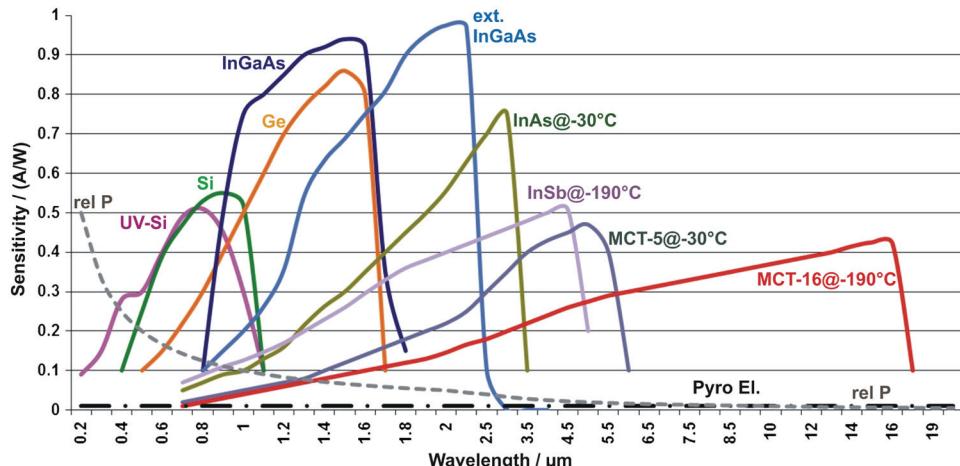


Figure 5.6 Simplified efficiency curves of solid state detectors for the range $<20 \mu\text{m}$.

detector spans, the lower its sensitivity tends while cooling improves the sensitivity besides the SNR. The curves shown are typical; real data depend on the material mix.

5.6.1 General effect of cooling

Diode detectors deliver a signal that is linear with respect to intensity. A dedicated pre-amplifier converts the charge or current into a voltage. The efficiencies of semiconductor materials reach up to 0.9, but they also produce high thermal background. For example, the efficiency of a Si detector will beat a PMT. At high signal levels, that will lead to a better SNR. In low-level situations, in turn, the PMT provides far better results. That difference is documented by the detection limit and SNR curves. The thermal behavior of the different solid state systems is not equal but similar. A rule of thumb applies: each reduction of 5–10 K cuts the background in half. Cooling an element to -25°C , or 45 K below ambient, a background reduction of a factor of 50–100 can be predicted. Cooling by liquid nitrogen (LN, -196°C) reduces the background to very low signals. It makes no difference whether the detector is of the photovoltaic or of the photoconductive type, the thermal background reduces the available linear range, increases the background signal, and drives up the noise. In summary, the thermal environment of a solid state detector has a strong influence on the limiting SNR.

5.6.2 Planck's radiation equals blackbody radiation

The spectral distribution of different temperatures can be simulated by Planck's law of blackbody radiation. The luminance of thermal radiation, multiplied by the spectral efficiency of the detector, creates measurable values already at “low” temperatures.

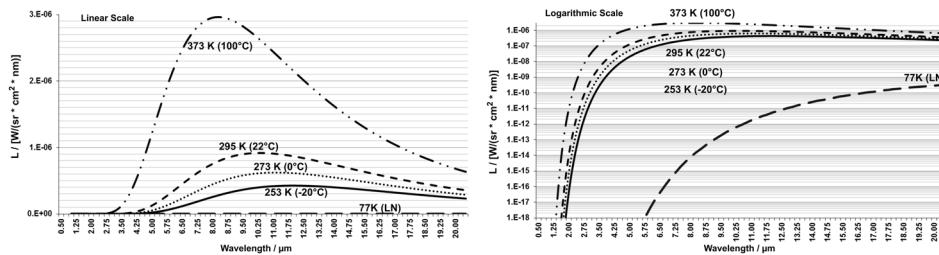


Figure 5.7 Planck's radiation at low temperatures in linear and logarithmic scale, in the range of 0.5–20 μm .

In Fig. 5.7, the curves represent temperatures between -20°C and 100°C . All of them produce reasonable and increasing signals from between 2 and 20 μm in wavelength. The lower dashed curve is valid for LN at 77 K, or -196°C . It is too minor to be visible in the left graph.

Planck's radiation appears as a background signal on one hand, and on the other as a desired signal, as described later in Chapter 6. The background radiation is a signal that adds to the desired signal information as far as it appears in the detector's wavelength range. The room temperature of 22°C is the second strong-dashed curve in the figure. Depending on the detector's response, it might affect operation from 800 nm upward, as already mentioned in Section 5.5.

5.6.3 Detectors and the ambient temperature

The background charge's impact strongly depends on the detector material. Generally speaking, the magnitude increases exponentially with the IR wavelength and in a square function with the bandwidth. A few room-temperature numbers are provided for comparison. The background of a standard silicon detector element (350–1150 nm) with a 1-mm diameter is set to be the reference (magnitude 1). A UV-Si element (180–1150 nm) will produce a magnitude of 1.4. Even at room temperature, it creates no reasonable signal below 350 nm, the small bit and the increased bandwidth combined produce a higher background. A 1-mm InGaAs-1700 (700–1700 nm) creates $\sim 30\times$ the background. InGaAs-2200 provides a factor of 100, and InGaAs-2600 a factor of 300, compared to the reference. These are good reasons not to extend the upper wavelength limit wider than required. An In-Sb will carry a thousand-fold background. In general, the background steeply increases farther into the IR “looking” detectors. The background also increases with the square root of the interface volume of semiconductor detectors. A 10 mm \times 10 mm silicon detector will produce a background signal of 10^4 compared to the 1 mm \times 1 mm element. Of course, spectroscopists want large detectors to collect all light leaving the monochromator. The wider the wavelength reaches into the IR, the harder it becomes to find the right balance between size and background. Additional problems are price and availability of elements larger than 3 mm \times 3 mm. When

it comes to SNR and background, frequency-modulation techniques are the solution to the dilemma discussed next. The study requires the review of frequency-dependent parameters such as border frequency, damping, and rise/fall behavior, as presented in Section 1.7.

The parameter D^* combines important specifications of IR detectors in a single equation. D^* employs $(\text{cm} \times f^{1/2}/\text{W})$ for its result. The size is related to 1 cm^2 , the modulation frequency to 1 Hz, and the entrance optical power is normalized to 1 W. The resulting SNR defines D^* , which allows comparison of all detectors. The higher the number is, the better the detection limit of the device under consideration. That behavior makes D^* a good piece of information in data sheets. On the other hand, if one wants to compare detectors for a given experiment, the following equation might be better:

$$D^* = (\text{SNR} \times \Delta f^{1/2}) / (P \times A^{1/2}), \quad (5.2)$$

where D^* represents the detection capability, SNR is the signal-to-noise ratio, $\Delta f^{1/2}$ is the square root of the electrical bandwidth, P is the optical irradiation (in watts), and $A^{1/2}$ is the square root of the area of the detector element (in cm^2). The applied frequency and its bandwidth are two important IR parameters, which are treated in the next section.

Figure 5.7 notes that the temperature of the detector element has a strong impact on D^* . Two more circumstances are relevant. Even if the element itself is cooled, say to -30°C , the element will recognize the temperature of the housing, windows, and other hardware, and add the information to the spectroscopic result. To minimize that factor, detectors with LN cooling often include a “cold aperture,” which is a mask that limits the viewing angle of the element so that only the required angle for the spectrometer exit is left. The cold aperture is also cooled by LN, which makes the method efficient, especially if the aperture is designed for the actual spectrometer system.

In Fig. 5.8, the viewing angle of the detector element is limited by the cold faceplate. The element and faceplate are connected to the Dewar volume by a thermal bridge (T guide). A hole in the faceplate will, in the best case, define

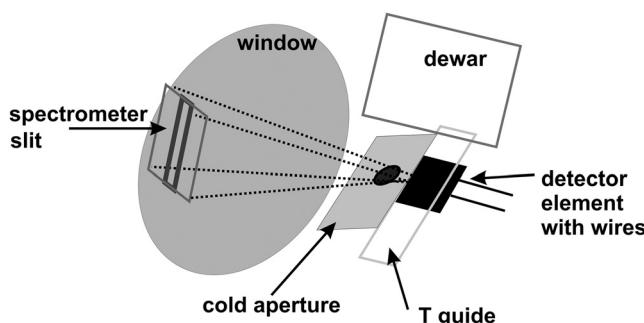


Figure 5.8 Principle of a detector with cold aperture.

an aperture angle optimized to allow just the spectrometer slit to be viewed. The window of the housing can include a lens or filter function. Thus, the detector will only “look” at the warm slit and a little bit into the also-warm spectrometer to receive the thermal background. The temperatures there are perhaps 20°C or even more. It would require much effort to fight that behavior, such as using an evacuated spectrometer and a cool focusing mirror and output slit. That would shift the background down even further, but at what expense?

5.6.3.1 Signal modulation and synchronized detection

Consider the following comparison of steady-state data acquisition and modulated signal gathering. Two exemplary simulations will demonstrate the SNR of IR detectors. The task is to perform an absorption measurement in the wavelength range of 3.5–4.5 μm with a 1-nm bandwidth. The spectrometer used will need a slit width of 1 mm; the slit can be used up to a 4-mm height. The maximum slit width is 4 mm, and the viewing angle of the detector will allow for a 4 mm × 4 mm aperture. The detector is equipped with an indium–antimonide element with a 4-mm diameter and an area of 0.16 cm². It is LN cooled (77 K) and incorporates a cold aperture. The D^* is between 10^{10} and 10^{11} ; the efficiency grows, as shown in Fig. 5.6 (the pink curve “InSb”), from 0.48 A/W at 3.5 μm up to 0.51 A/W at 4.5 μm. The built-in current/voltage converter can be set to 300-, 3,000-, and 30,000-fold. The output will reach saturation at 10 V and the element itself at 12 mA, which equals an optical entrance power of 24 mW. The detector housing is equipped with a 1:1 refocusing, cooled lens, which must not be taken into calculation. The element observes a cone of 1 steradian (sr). The light source is a 50-W Globar with a 1000-K temperature. The ambient temperature is 22 °C, equal to 295 K.

The calculation of the detector is performed as follows. At 77 K, a power-equivalent background (in Fig. 5.7, the lower dashed curve) between 4.5×10^{-15} W/(cm² × nm) at 0.7 μm and 1.32×10^{-16} W/(cm² × nm) at 5.2 μm is created, which produces an integral of 1.666×10^{-16} W/(cm² × nm), or 0.16 fW at an area of 0.16 cm². Because the background almost completely comes from the wavelength range around 5 μm, an efficiency of 0.5 A/W is multiplied by the maximum gain of 30,000. The result: the detector itself creates an output signal of 2.5 pV (rounded) within the 10-V output range. Theoretically, a dynamic range of 4×10^{11} is allowed.

Note the impact of Planck’s background irradiation from the environment of 22 °C, the second strong-dashed curve in Fig. 5.7. Between 0.7–5.2 μm (the working range of the InSb), it carries a background irradiation between 3×10^{-27} W and 2.3×10^{-7} W; the integral sums up to 0.185×10^{-3} W/(cm² × sr × nm). The observed area (at the output slit of the spectrometer) is also 0.16 cm² at 1 sr. The background is converted at an average efficiency of 0.5 A/W and a gain of 30,000, creating an output signal of 1.4 V (rounded). That makes the detector’s own background negligible by far.

The sole utility signal with a 1-nm bandwidth can now be considered. The source, at a wavelength of $3.5\text{ }\mu\text{m}$, irradiates $0.936\text{ mW}/(\text{sr} \times \text{cm}^2 \times \text{nm})$. The optical power constantly declines to $0.57\text{ mW}/(\text{sr} \times \text{cm}^2 \times \text{nm})$ at $4.5\text{ }\mu\text{m}$. Furthermore, the losses in coupling and the experiment itself must be accounted for. Suppose that 25% of the power reaches the detector, given a median source irradiance of 0.7 mW , which will deliver 0.175 mW to a surface of 0.16 cm^2 and 1 sr, or 0.226 mW at the active surface of the detector. This value is to be multiplied with the average detector efficiency of 0.48 A/W and the electronic gain of 30,000. Overall, a wanted signal of 1.693 V at $3.5\text{ }\mu\text{m}$ is produced, which drops to 1.03 V at $4.5\text{ }\mu\text{m}$. In other words, a signal/background ratio of approximately 1:1 and a similar SNR can be expected. This realistic scenario will not produce results better than 7:1 in signal/background ratio and 3:1 in SNR. Reasonable absorption measurements are not possible. Because the detection limit is also defined by the background, emission measurements are only useful if the output signal exceeds 1.4 V , which is why experiments with continuous light are rarely performed in the IR range. Multi-frequency modulation systems with Fourier transform analysis (which are not the topic of this work) perform much better, and the modulation is conducted at a fixed single frequency and with synchronized detection, as is the case with lock-in amplifier technology.

The technology and parameters of lock-in amplifiers are supported by accessible literature, so the discussion here is limited to the facts important for spectroscopy. The signal is modulated at a certain frequency, and the phase relation of the signals is known and fixed between the modulator and synchronized amplifier (locked in). The detection includes the magnitude and the phase relation, as well; the signal is detected for its real and imaginary parts. This process is performed by phase sensitive detection (PSD), in which the signals from the modulator (the reference) and those from the detector (the sample) experience a complex multiplication. The arithmetic leads to the fact that only signals with the same frequency and phase will, in the long term, contribute to the output signal. All deviating signals integrate towards zero over time. The modulation is achieved by controlling the light source or chopping the light. (Both methods are compared in Chapter 6.) This comparison uses a chopper and locates it between the DC light source and the experiment. It is assumed that the In–Sb detector is designed for DC to 3 kHz. The higher the frequency is, the better the filter function Q , where $Q = \text{frequency}/\text{bandwidth}$. The Q reduces the noise linearly, and the SNR improves by the square root of Q . However, a chopper does not work equally well at all frequencies and creates increasing acoustic noise. The working frequency must be kept below the corner frequency of the detector, here 3 kHz [see Eq. (1.13)]. At 3 kHz, the signal from the detector will already be 3 dB, or $\sim 14\%$, below maximum, and the phase will be shifted, too. The reasonable maximum frequency of a detector is $\sim 2/3$ of the corner frequency, equaling 2 kHz in this case. On the other hand, lowering the frequency too

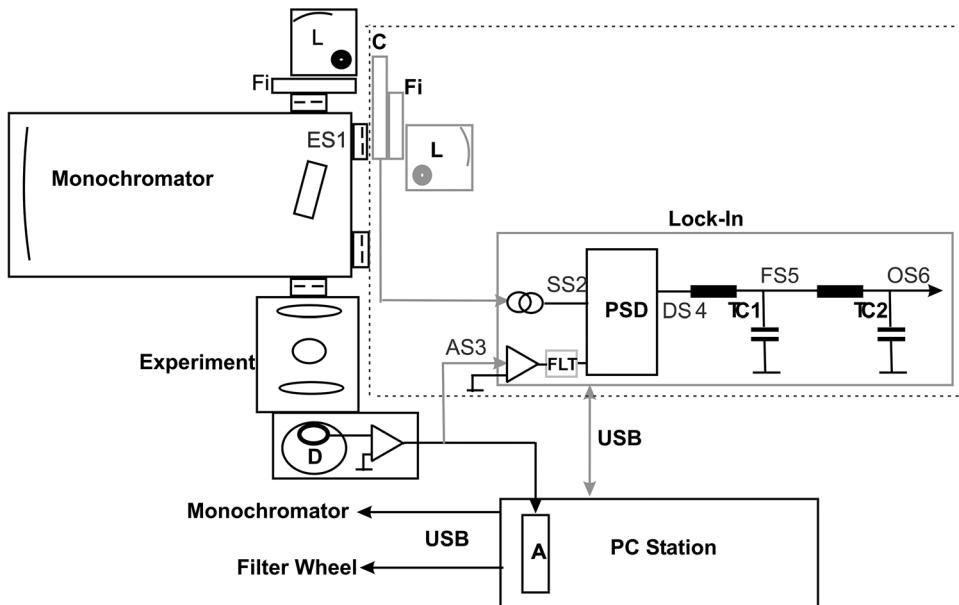


Figure 5.9 A DC system, shown in black; the additional components for synchronized detection are marked grey and framed.

much will open the door for fluctuations and drift, known as $1/f$ noise. A good value for chopped spectroscopy applications is a range between 80–300 Hz. To avoid harmonics and technical frequencies, and because the example chopper runs easily and quietly, 222 Hz is selected for this scenario. More information on bandwidth, e -function, and attenuation can be found in Section 1.7.

The typical lock-in signals, marked grey, are shown in Fig. 5.10. The black components indicate the signal path without lock-in: the light source (L) focuses on the monochromator entrance slit. The light passes an order-sorting filter wheel (Fi). The monochromator, its periphery, and the lock-in are controlled by a PC. The dispersed light passes through the sample and hits the detector D. In the calculated example, it is equipped with an LN Dewar for optimal behavior, which incorporates a dedicated current/voltage converter with variable gain. The output signal is led to the PC with an analog/digital converter board (A) to read and process the data.

The grey-framed components show that the signal path with lock-in differs in one position. Between light source and monochromator entrance, a light chopper (C) is placed. That also creates a steering signal to synchronize the lock-in. This signal is routed to the reference entrance and controls the PSD. The detector signal is led to the analog entrance of the lock-in, amplified (optionally filtered, FLT), and analyzed by the PSD. After final filtering it is guided to the PC, stored, and processed.

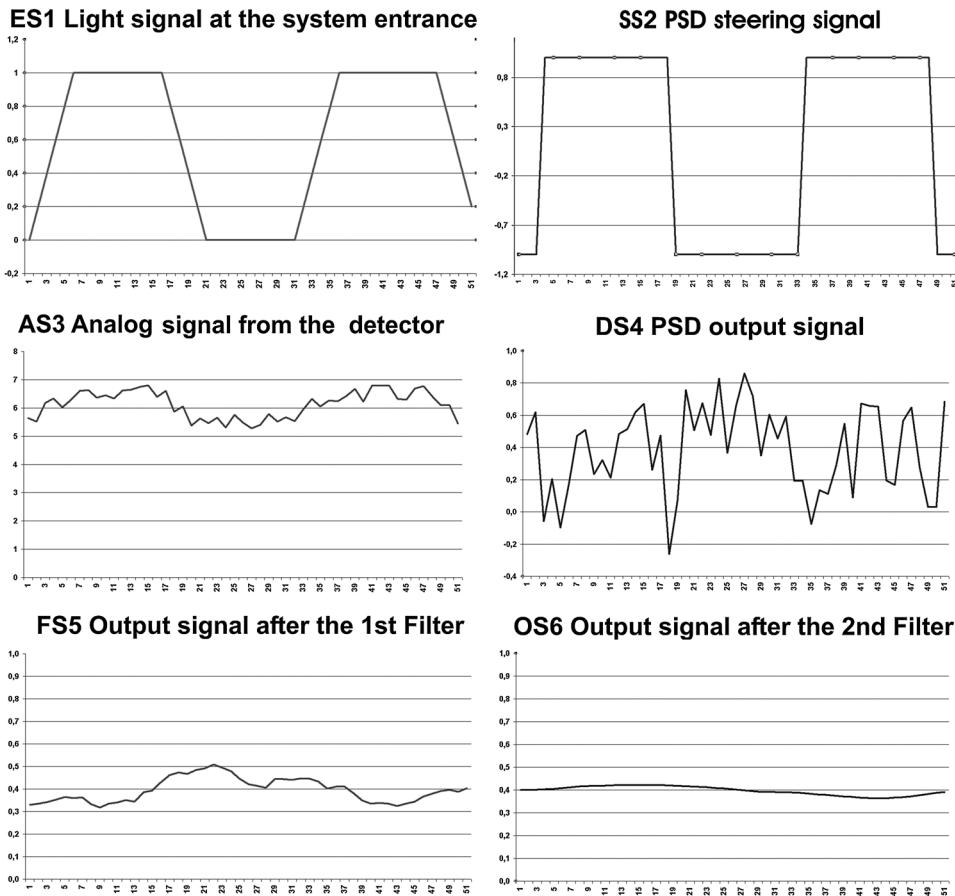


Figure 5.10 The lock-in amplifier's most important signals.

Typical signals at the lock-in of optical applications

The chopper delivers a perfect rectangular reference signal. Polarity is between 0 and +1, and is typically TTL-compatible.

- **ES1:** The chopped light often does not have a perfect rectangular shape because the expansion of the light spot tilts the flanks.
- **SS2:** The lock-in incorporates signal-shaping circuitry to ensure that the reference signal is either a sine wave or a perfectly rectangular one. That signal will control the PSD and is of symmetric polarity, between -1 and +1. To allow for the compensation of differences in travel time and to overlay both the control and measurement signal, the circuitry also contains a phase-shifter. To allow accurate alignment, the phase position is shown and can be monitored by an oscilloscope. If the control signal is a sine, only the portions of the basic frequency (first harmonic) contribute to the integrated output

signal. If the control signal is rectangular, the odd harmonics will also contribute. The latter is of advantage to optical spectroscopy with chopped light because the signal can increase by up to 22%.

- **AS3:** The detector's output contains the utility signal, which is modulated and of a defined phase, but the signals ride on a thermal DC background. In our example, the ratio is about 1:5. The lock-in entrance will block all frequencies below a certain limit. In the more flexible models, the limiting frequency can be chosen (in this case, ca. 20 Hz would be a good choice). If it is not selectable, the limit in most cases is in the range of 0.1–10 Hz. The pre-amplifier might offer additional filter functions. The best models can read voltage, current, and charge, and be switched between grounded or earth-free (free flowing). If that is the case, the signal can be sensed at the detector element directly.
- **DS4:** This part is the complex product from control and measurement signals. After the PSD, the signal parts of correct phase relations will be integrated, because they are of complex synchronicity. Because the PSD only works with AC signals, its entrance signals are bipolar, which means that they vary between $-x$ and $+x$, and $-y$ and $+y$. If both are positive, the output product is positive, too. If both are negative, the product is positive, as well. If they are perfectly opposite, the output is zero. If they not fully synchronous, the product tends towards zero, and that is true for all kinds of noise and interference. The complex multiplication effects that the negative signal shares will be “turned up” and contribute to the integration. A lock-in always works with virtually symmetric signals, even if, as in the example presented here, the detector provides a unipolar signal. Comfortable “dual-phase lock-ins” supply two PSDs that can directly compute magnitude and phase. The benefit is that phase position and drift play no role. However, the magnitude is calculated from the square root of the squared real and imaginary parts, and thus no negative data are possible, which is counterproductive for noise analysis and reduction because the average will never be zero. Therefore, it is better for the data to run the basic mode “real and imaginary detection” for the lock-in and take care of the phase relation.
- **FS5:** The product of the PSD is phase-critical and complexly multiplied, but not yet integrated. Thus, it varies with the modulation frequency. A simple low-pass, made by a resistor in series and a capacitor to ground, creates the time constant τ and affects integration, as shown in Section 1.7.1. τ should be at least as long as the time required for five signal cycles to avoid interference and switching signals.
- **OS6:** To flatten out the output signal even better but allow for fast signal-response times, two identical RC filters are often put in series. The damping (“roll-off”) then increases from 6 dB/octave to 12 dB/octave. See Fig. 1.6 for the difference.

Table 5.1 Signals created by the DC and lock-in techniques.

Method	Time Efficiency	Signal Efficiency	Output Signal Magnitude	SNR	Time Factor Required to Reach SNR = 1
DC (stationary)	1.000	1.000	1.000	1.000	1.000
DC (clocked)	0.800	1.000	0.800	0.894	1.250
Lock-in	0.500	0.800	0.400	0.632	2.000

Table 5.1 contains some parameters of three data acquisition modes. “DC stationary” describes a static system that does not take any background during data acquisition. “DC clocked” is a system that takes a new background after every tenth data point. The actual background is subtracted from the next 10 data. The lock-in method also assumes a continuous light source, and modulation is done by a chopper. The SNR is the product of the two efficiencies, whereas the required measurement time for equal SNR comes from the squared ratio of time and SNR. The fact that the lock-in method cancels out disturbances that the other two methods will integrate into the output signal is not considered here. The lock-in method indeed provides better fidelity.

If a DC light source is used, and the signal from the detector is acquired continuously, the output signal represents the mean value. The DC stationary method represents a system taking one background before the data gathering starts and reducing that single background from all data points. The DC clocked method is a time-controlled system that takes a new background after every n^{th} data point and reduces that from the following data. In the table, $n = 10$. It is assumed that a fast mechanical shutter does not need more time than the background integration itself. Many systems will not incorporate a shutter that fast; therefore, the beam transfer efficiency versus time will be less. The lock-in method only uses 50% of the DC light, leading to half the time efficiency. The chopper will create a signal flank of 10% (as in ES1 of Fig. 5.10), wasting another part of the light. In the end, only 40% will be integrated. Because the measurement time for equal SNRs follows a square function, the time required for the same SNR increases faster than the signal drops.

Several routines help optimize the stationary method. One way is to gather backgrounds before and after acquisition. Thus, a possible change can be recorded, and the mean value can be reduced, if required. Regardless, all synchronous and asynchronous noise will contribute to the output signal. The clocked method provides the advantage of fitting the repetition of background acquisition to the drift requirements. Eventually, the background of both methods is stored in a separate file for later evaluation and processing. Again, all synchronous and asynchronous noise will contribute to the final signal.

The lock-in method is the slowest of the three. Its advantage is fully automatic background compensation combined with a strong filter function.

Depending on the parameters of the bandwidth (damping) and entrance filter used, it is possible to let only signals pass that are synchronous to the chopper's frequency and phase. If the background and desired signal are close, as could be the case in IR applications, the lock-in method (besides FFT techniques) might be the only means of safely acquiring useful data.

The chopper in the example presented at the beginning of this section runs at $f = 222 \times s^{-1}$, and the inversion $t = 1 \text{ s}/222$, or 4.5 ms. The angular frequency is important for the filter function of the circuit because it defines the bandwidth. According to Eq. (1.13), $\omega = 1395 \text{ Hz}$. The detector has a bandwidth of 3 kHz. The numbers show that there is almost no filtering from the modulation itself. The lock-in's advantage comes from the PSD, the signal integration, and the damping (frequency attenuation). The variable RC circuit right after the PSD, with variable time constant τ , responds e -functional and requires $\sim 5 \tau$ to reach 99.33% after a change in the system input. Often, 5τ is called the “response time.” Supposing that the time constant τ of the lock-in amplifier is set to 1 s, the response time will be 5 s. That behavior lets the amplifier work like a band filter. The bandwidth is identical to the angular frequency of 6.28 Hz. Signals at that distance from the center frequency, even those within the bandwidth, are already attenuated by 3 dB (to $\sim 70\%$). The quality of the measurement can then be calculated: $Q = 222 \text{ Hz}/6.28 \text{ Hz}$, $Q = 35$. If $\tau = 1 \text{ s}$ is not useful and it must be, i.e., 0.1 s, the filter quality would be only 1/10, resulting in $Q = 3.5$, which would allow a response time ten times as fast. The filter function, on the other hand, can be improved by increasing the Q of the lock-in entrance; that can be achieved via additional entrance filtering. Some lock-ins incorporate such filters that are selectable for several functions. For instance, a bandfilter with a 10-Hz bandwidth at the entrance will produce $Q = 222 \text{ Hz}/10 \text{ Hz} = 22$ and increase the total $Q = 22 + 3.5 = 25.5$, while the response time stays at 0.5 s. The output filter will not suffer from any drift because the frequency and PSD are always locked to the reference. The shortest technically useful time constant is five times the cycle time of the modulation frequency; in the example $\tau = 22.5 \text{ ms}$, 25 ms might be a value of choice. If too small a τ is chosen, variations and interference in the output signal of the lock-in will occur resulting from incomplete integration of real and imaginary parts of the demodulated signals. The SNR of a system with lock-in is already at a minimal time constant, far better than the results achieved by a DC mode. SNR increases with the square root of the increase of the time constant, i.e., a factor of 6.3 at an increase from 25 ms to 1000 ms. A practical time constant for scanning systems is 0.1–3.0 s.

5.6.3.2 Estimation of the modulated measurement

An estimation can be found using the conditions from the measurement with continuous light listed in the previous section. Likewise, the signal levels at the detector remain unchanged by the modulation. Thus, the background level will stay close to 1.4 V, caused by the 93 μW of background on which the

utility signal is “riding.” Under all circumstances, it is necessary to keep the sum of all signals within the linear detector range. The synchronized measurement creates new, frequency-dependent noise. The example detector discussed here works from DC to 3 kHz and has its own equivalent noise of $2 \text{ pA}/\sqrt{\text{Hz}}$. Thus, it adds an equivalent background of 2.83 pA, creating 42.5 μV at the output. All continuous light and the unmodulated background signals will then be suppressed. Thus, the noise background of the detector is also the total background and will be in the range of 4×10^{-6} of full scale. It is obvious that the noise of the detector itself has very little impact.

Now consider the utility signal. The Globar light source at 3.5 μm creates a power of $0.936 \text{ mW}/(\text{sr} \times \text{cm}^2 \times \text{nm})$, decreasing to $0.57 \text{ mW}/(\text{sr} \times \text{cm}^2 \times \text{nm})$ at 4.5 nm. Only half of the power is transferred to the monochromator (and thus to the experiment) because the chopper blocks half of the light. After the chopper, the losses are the same as before. The output is then roughly 0.85 V at a wavelength of 3.5 μm and at a 1-nm bandwidth, decreasing to 0.515 V at 4.5 μm . Without further amplification and filtering, the signal/background ratio is already 1.6×10^4 . That number is good for linear measurements up to 3 absorbance units, or 0.1% transmission. In addition, the detector’s noise background will be very stable and tend towards zero by the lock-in. The utility signal can be easily amplified to convenient values by variation of the lock-in gain; the SNR can be improved by additional filtering. Even without further optimization, the detection limit will be equivalent to the magnitude of the internal noise of $\sim 3 \text{ pA}$, or 6 pW at the detector. The application of the lock-in amplifier transforms the whole setup into a useful, quantitative spectroscopy system.

5.6.4 Tandem detectors

Silicon’s ability to become slowly transparent between 1120–1170 nm destined Si sensors to be used in tandem (“sandwich”) with other detector elements. The Si diode is mounted on top of the longer wavelength-sensitive element. Both will be read separately but cooled together, if required. This arrangement can achieve increased working ranges: for example, a combination of Si with In-Sb can reach from ~ 200 –5300 nm. Because the characteristics of both partners remain, remarkable differences in sensitivity, background, and noise can occur. The two separate output signals of the elements can be added at the expense of a noisier element dominating the level. The total efficiency curve of the tandem is the combination of the two materials.

As shown in Fig. 5.11, the silicon element is mounted on top of the IR element. Both have separate pre-amplifiers and outputs, completed by a summing output. If TE cooling is integrated, the Peltier element usually has the shape of a jar, enclosing the sandwich not only from the bottom but also from the sides. The total efficiency curve of the tandem is the combination of the two materials. The combination UV-Si (solid curve) with InGaAs-1700

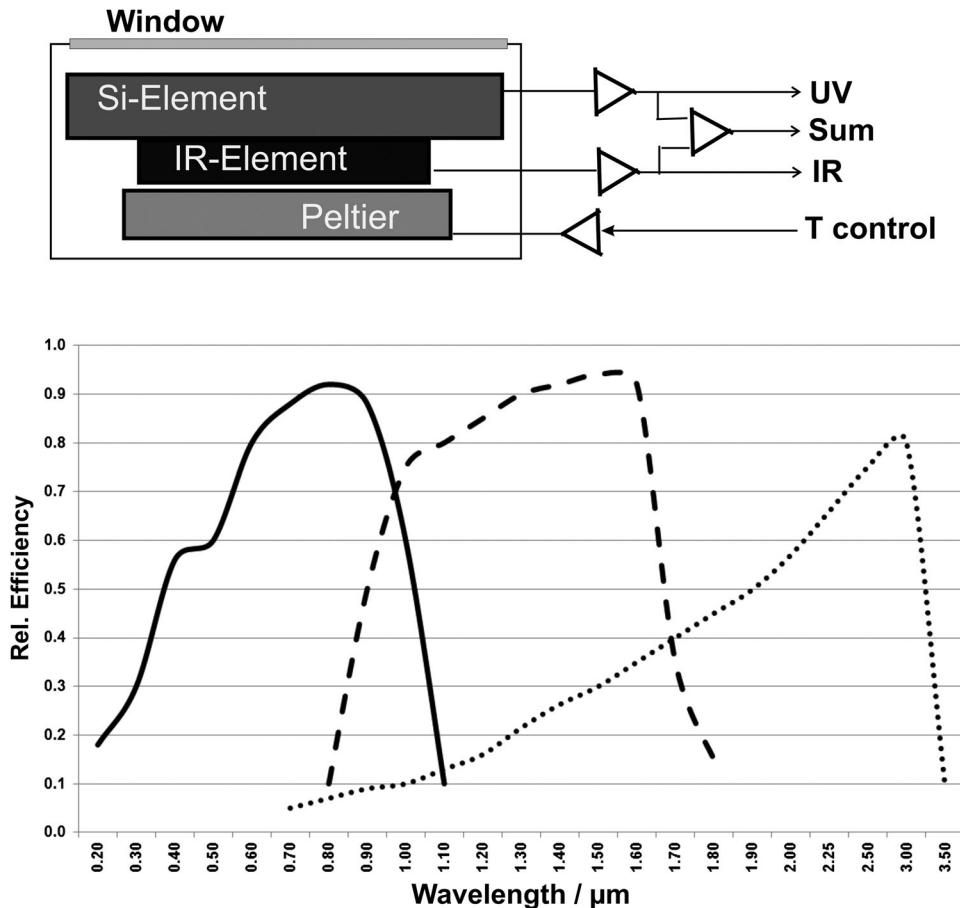


Figure 5.11 Principle of sandwich detectors.

(dashed curve) combines very well in the range of the boundary. Conversely, if UV-Si is combined with In-As (dotted curve), there will be a deep depression in sensitivity in the range of 1.0–1.5 μm . The SNR of the materials (D^* can be used for comparison) also differs strongly. A third problem might be the size, as Si elements with up to 10-mm diameters are no problem, whereas InGAs or InAs larger than 3 mm are difficult to obtain. A system cooled to -25°C , comprising a UV-Si detector with a 5-mm diameter, will have an internal noise of $10\text{--}30 \times 10^{-15} \text{ A}/\sqrt{\text{Hz}}$. A 3-mm diameter InGaAs-1700 will be at $\sim 100\text{--}200 \times 10^{-15} \text{ A}/\sqrt{\text{Hz}}$, but a 3-mm diameter In-Sb will create $\sim 10\text{--}50 \times 10^{-12} \text{ A}/\sqrt{\text{Hz}}$.

5.6.5 Typical parameters of solid state detectors, and their interpretation

- **Gain:** 10^6 V/A means that a photocurrent of $1 \mu\text{A}$ creates an output signal of 1 V .

- **Maximum output signal:** In most cases (1 or 10 V), possible saturation effects are not taken into consideration. If the used range is kept below 80%, enough overhead is provided for almost all cases.
- **Output noise:** This value, divided by the gain, produces the equivalent noise at the input. If the output noise is 1 mV and the gain is 10^6 V/A, the input noise is 1 nA.
- **Efficiency (A/W):** This value allows one to calculate the light input power and/or input photons. If an instrument has selectable frequency bands, each band requires an extra specification of noise.
- **Bandwidth:** This value defines in what modulation frequency range a detector will work. By the angular frequency, one can calculate the rise and fall times in response to rapid signal changes. Attention should be taken into account because the angular frequency ω , also known as f_c , marks the -3 dB (rounded to 86%) limit.
- **Frequency range:** If the lower frequency border is DC, the system can be used for continuous light. However, that implementation is more liable to drift. It is useful to be able to select between DC and AC, with the lower- and upper-corner frequency also selectable.

5.7 Design Considerations of Solid State Detectors

5.7.1 Illumination of small detector elements

Because the background increases with the square function of the element area of semiconductor sensors, small active areas are preferred (a 3-mm diameter is typical). If the spectrometer slit or the light cone (because of the slit-element distance) is larger than the element, transfer optics make sense. The examples shown in Figs. 5.4 and 5.12 can be applied here but in reverse in order to create a condensed image at the detector.

5.7.2 Charge storage in semiconductor elements, thermal recombination, and holding time

Sensors such as silicon, germanium, and InGaAs can also be manufactured to store and integrate the charge carriers, at which point they are no longer diodes but transistors. That application allows signal storage in the interface layer; there, the charge carriers are stored until the read-out, which includes discharge or recharge. The SNRs are similar to those of diodes. Because the carriers are stored for a defined time between two read processes, the recombination rate plays an important role. It is an exponential, thermal process that either reduces the signal or increases the background (which also means that the signal decreases). For silicon, the time constant of the thermal process at room temperature is in the range of ~ 300 ms; for InGaAs, it is ~ 150 ms. If a read-out process loads the interface fully, the charge after 1τ will

be reduced by 63%, and after 5τ , the charge will practically be gone. The thermal effect of recombination happens parallel to and in competition with the photon conversion. Thus, it is not possible to determine whether a signal is created by light or heat. It makes no significant difference if the read-out is charging or discharging, the result is similar. The good news is that the thermal background is reproducible, as is the noise it carries. If read frequently, e.g., every 10 ms, the background will stay $<1\%$, superimposed by noise of $\sim 2\%$ of the maximum signals. If that is sufficient, no cooling is required. Because the recombination rate of the elements is defined as 1τ , it must be considered that after one time constant 63% of the capacity is filled with thermally created charge carriers. Cooling effects are identical for transistors and diodes. The recombination rate of all kinds of elements (diodes, transistors, arrays, CCDs, etc.) will be halved per -6 K to -8 K . At $-25 \text{ }^\circ\text{C}$, the holding time is ~ 150 -fold compared to room temperature. For silicon, that scenario results in a time of $\sim 50 \text{ s}$ for 1τ . LN cooling will affect the recombination rate such that it allows signal integration over many hours or even days.

5.7.3 PIN and avalanche diodes

In a normal detector diode, the positive and the negative poles are separated by the interface, and the sequence of electrodes is P:N. If a nondoped layer, an intrinsic layer, is added in between the two, the specific field strength drops and so do the background and noise (the capacity also decreases, unfortunately). Such a detector is called a PIN (positive–intrinsic–negative). The technology is only available for small diodes with a small surface, typically a 0.5–1.5-mm diameter. Because of the low capacity, PIN diodes are used for small slit apertures and low photon currents.

The avalanche diode is another special version. The interface layer is thicker, and high acceleration voltages are sent across it. A detected photon creates one or more electrons that will be accelerated and create secondary charge carriers, even within the interface. The result is a cloud of charge carriers (avalanche) at the pre-amplifier for each detected photon, which suits avalanche diodes for counting systems. They are not useful for analog use because with each pulse the amplification voltage collapses inside the layer. The recovery time defines the maximum counting rate, which is typically below 100 ns, allowing count rates of at least 3×10^5 events per second (pps).

5.7.4 Detector coupling by fiber optics

As shown in Fig. 5.12, one method of coupling uses optical fibers. For some kinds of detectors, this method has additional advantages. Instead of a window, the enclosure can be realized by a brush of optical fibers. That, in turn, allows the coupling of a fiber cable with almost no losses.

The figure demonstrates the flexibility of fiber optic transfer with a monochromator versus a spectrograph. On the left, the cable receives the signals after the output slit and guides them to a circular detector. On the

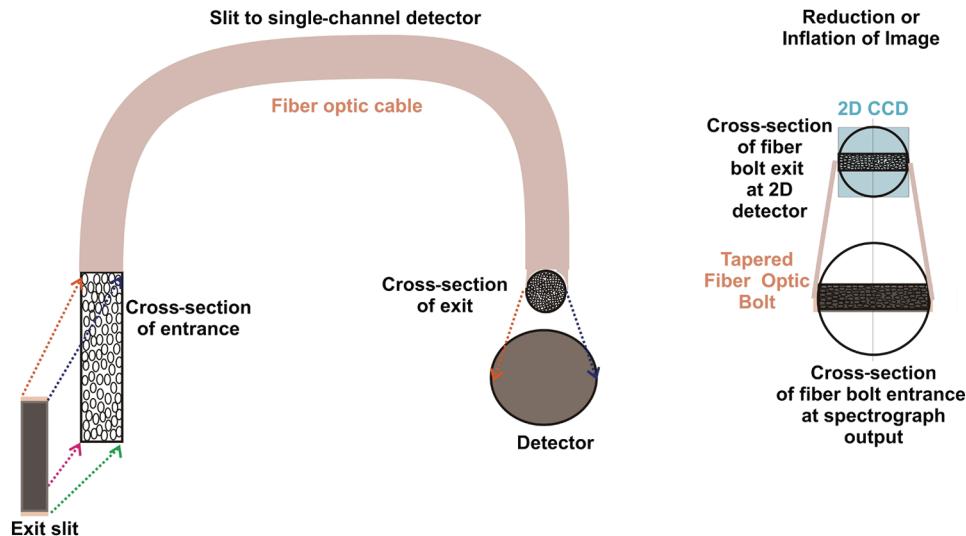


Figure 5.12 Detector coupling by fiber cable and fiber taper.

right, a fiber bolt is shown placed in the focal plane of an imaging spectrograph, transferring the signals to a smaller 2D detector. In both cases, as much light as possible is collected. Several methods of coupling can be realized with a certain length of fiber cable or bolt, provided that the shape can be modified, such as conversion from disk to rectangle, or inverted, as shown. The left example modifies both the shape and size via a fiber cable, whereas the right part shows reduction by a solid bolt. The modification is realized when each fiber is pulled during production and before assembling; it does change the cross-section at the side of pulling for each single fiber. Fiber brushes and cables that incorporate a change in the cross-section or shape are called “tapers.” In addition, the distance can be used to fit the cone of light and active area at both ends. The losses in a dual-side, “air-coupled” system are in the range of 25%. By utilizing brush coupling and applying an optimal coupling medium that consists of the same refractive index as the fiber material, the loss is reduced to very small percentage. Index-matching oils are available to accomplish the match. A further benefit is the avoidance of aberrations. Tapered couplings are often found in combination with CCDs, arrays, or image intensifiers, such as MCP, as discussed in the next section.

5.8 Area Detectors: CCDs and Arrays

5.8.1 Mounting of area detectors, the resulting disturbance, and the distribution of wavelengths

Diode arrays, CCDs, CIDs, CMOSs, and other detectors mounted in the focal plane of spectrographs are thoroughly documented in scientific literature and by manufacturers; thus, the specifics relevant to spectroscopy systems are discussed

here. Consider the following negative argument: A classical spectrometer does not want to receive light from the output, and it does not want to refocus different parallel signals onto a flat area. It wants to focus on one radius and in one orientation only (the spectral one). The magnitude of aberrations from the ideal situation that need to be taken into account is discussed in Sections 2.6 and 4.1. Falsifications in the beam travel (“corrections”) improve the imaging behavior at the expense of the pure monochromator performance. This behavior should be kept in mind when deciding on the purchase and application of flexible spectrometers. In this book, spectrometers not designed for general use and for different detectors are not the topic. Instruments limited to a single mode of operation or a single application can provide excellent optimization, whereas flexible instruments must compromise.

Unfortunately, all known area detectors are built in such a way that the housing’s window will reflect back into the spectrometer, which increases stray light. Thus, an AR coating is beneficial. Furthermore, because the window is perpendicular to the almost-parallel beam, interference can occur and overlay the utility signal. Wedged windows are available, and some manufacturer will, on request, place the window at an angle. Both variations reduce the reflection problems, but they are not always listed in the product description, so it might be necessary to ask.

The sensors are also perfectly mounted in the focal plane (for valid reasons), which also can lead to interference and back-reflections. Some sensors can be furnished with an AR coating. Many sensors have their own window in addition to that of the housing, which is another source of interference and losses, especially if the distance between the two windows is short. The leading manufacturers of scientific CCD cameras deliver (again by request) systems with only one window. Some detector adapter flanges of flexible spectrometers allow slight tilting in the vertical. Even one degree of tilt can help reduce interference if the imaging or spectral performance can tolerate it. Section 4.15 states that the local dispersion (in the plane of the detector) depends on the angles at the grating, which, in turn, depends on the wavelength and focal length. No two output positions will have the absolute same dispersion. The variation should be taken into account and, if possible, calibrated out. The solutions require third- or fourth-degree polynomial functions. Useful camera software will have the necessary tools to correct variation either by numerical or online calibration, in the correct dimension of wavelength/wavenumber/eV, and store the correction parameters. The narrower the angles of output light are, the longer the focal length of the spectrometer, the smaller the size of the detector, and the smaller the deviations detected.

5.8.1.1 Popular versions of area detectors

The most popular detectors used in the distribution of systems of laboratories are flexible, modular systems equipped with CCDs. Also frequently used are diode arrays, CMOS, and CID area detectors.

- **CCD:** A chain or 2D matrix of charge-coupled devices in MOSFET technology makes up an optical detector. The charge created in a certain pixel by the detected light is transferred to the amplifier electronics by the bucket-chain principle. The surface fill factor of a CCD is 1, but because very thin control lines are placed between the elements, front-illuminated CCDs will not receive all of the light. The efficiency can reach up to 0.9. With adequate cooling and a slow read-out, both background and noise are kept very low.
- **Diode Array:** A line of diodes is read out by field effect transistors (FETs). Between the diodes are gaps typically the size of the diodes. The gaps are not blind: the charge carriers created there will be shared between the neighbor diodes. The fill factor is only 0.5, but no light is lost. The efficiency can reach up to 0.9. The background and noise are reasonably higher compared to a CCD, but the capacity per pixel element is also higher.

5.8.2 Basic parameters of arrays and CCDs with and without cooling

The discrete elements of an array or a CCD are called “pixels,” a word derived from “PICTure EElement.” They have a very small surface per element, contrary to single-element detectors. CCDs, which are useful for spectroscopy, comprise element areas from about $10 \mu\text{m} \times 10 \mu\text{m}$ up to $50 \mu\text{m} \times 50 \mu\text{m}$. Diode array pixels are larger, typically $25 \mu\text{m} \times 2.5 \text{ mm}$. The linear dynamic range for spectroscopy applications will be at least 5,000, or 12 bits, which will be sufficient for measurements that compare A with B, such as quality control or on-line tracking. Ambitious measurements require a linear range of $>10^4$ (14 bits), whereas relative measurements would be best at a dynamic range of 10^5 (17 bits). Actual standard instruments offer a 12–16-bit dynamic range for a single-pixel read-out without further accumulation in memory. In comparison with single-element detectors, these almost always depend on the external signal integration, storage, and post-processing, which enables the collection of huge numbers of counts for a single data point. In the case of area detectors, the main integrator is the detector itself. The capacity is limited by the technical volume of charge carriers per pixel and is related to the background and read-out noise. Table 5.2 presents some examples, at the end of which is the most popular diode array. All parameters are valid for the wavelengths between 200–1100 nm.

In the parameter line, “e⁻” means electrons. Typically, a temperature of -120°C is reached at the sensor if LN cooling is applied. The ADC time is valid per process. The parallel and the read-out shift times are described in detail in Section 5.8.2.3.

Two facts are obvious: the capacity, and with it the available dynamic/linear range, increases almost linearly with the pixel size. The dark charge created per time unit strongly depends on the temperature. It is halved at every temperature reduction between -6 K to -8 K . The strongest sources of noise are the read-out electronics. Because the frequency bandwidth influences the noise, it grows nearly with the square root of the bandwidth, which in turn grows linear with the inversion of the read-out time ($f = 1/t$). Diode arrays and CCDs differ in the

method of charge carrier storage. After read-out, a CCD pixel has no charge, no free charge carriers. Detected photons create the charge and load the interface layer until either the capacity limit is reached or the read-out takes place. Thus, CCDs are perfect for absolute measurements. At very small signal levels, limited by the background noise, the interface noise is kept small because the charge is small. Arrays work inversely. At the read-out, the interface is fully charged. Arriving photons unload the charge in linear fashion. Thus, the interface has the highest charge before a signal is detected, superimposed by the highest noise. That, in turn, better suits array technology for relative measurements rather than absolute ones. Near 100% signal, the interface is largely neutral and carries the minimum noise. That scenario is good for relative measurements of small signal deviations, as in absorbance and reflection applications.

5.8.2.1 Pixel size, capacity, sources of noise, dynamic range, shift times, read-out time, and ADC conversion time

As the pixel area increases, so too does the capacity, which Table 5.2 demonstrates. Pixel sizes smaller than $10 \mu\text{m} \times 10 \mu\text{m}$ do not offer enough

Table 5.2 CCD parameter overview.

Pixel Size [μm]	Capacity [e^-]	Ambient Dark Charge [e^-/s] at $+25^\circ\text{C}$	Cooled Dark Charge [e^-/s] at -25°C	ADC Rate [μs]	Read-Out Noise [e^-]	Linear Dynamic Range	Parallel Shift [μs]	Read-Out Shift [μs]
5×5	1.6×10^4	0.3	0.04	2	6	2,600	2.0	0.02
				0.05	15	800		
9×9	8×10^4	0.4	0.06	10	5	15,000	0.3	0.10
				1	20	4,000		
		at -50°C	at -120°C					
13×13	12×10^4	0.5	0.10	10	4	30,000	30.0	0.10
				1	12	10,000		
20×20	20×10^4	1.0	0.25	10	4	50,000	15.0	0.20
				1	12	16,000		
24×24	27×10^4	1.3	0.30	10	4	65,000	8.0	0.20
				1	12	22,000		
26×26	30×10^4	1.0	0.35	10	5	60,000	18.0	0.15
				1	15	20,000		
$25 \times 2,500$	80×12^6	15.0	5.00	10	800	100,000	—	—
				1	2,000	40,000		

capacity for serious spectroscopy. Referring to the Rayleigh diffraction limit [Eq. (2.23)] reveals that very small pixel sizes (below $\sim 5 \mu\text{m}$) are useless. Supposing a slit width of twice the pixel width to utilize best possible operation, the entrance slit might drop below the Rayleigh limit, and interference phenomena and ghost peaks might appear. The pixel capacity must always be seen in connection with the read-out noise; the ratio between the two defines the final applicable linear range (the dynamic ratio). The read-out noise increases with the bandwidth of the ADC and with the square root of the ADC frequency. Thus, do not read out faster than is actually required in spectroscopy applications. A time factor, sometimes forgotten in the calculation of CCD read-out times, is the shift time in the vertical direction (also called parallel shift), which cannot be estimated from the size of a CCD or its pixels (it is a technical feature). There are CCDs with rather small pixels and slow shift time, and some with rather large pixels but fast shifting; it depends on the architecture. The cross-section of the control lines and their guidance plays a role. Regardless, well-designed cameras will offer access to the vertical (parallel) shift time, which allows the user to balance time versus linearity. Under normal use, the parallel shift time is not dependent on pixel grouping/binning. It is calculated by the time per shift times the number of lines of the CCD. The read-out towards the ADC (horizontal shift) also needs some time, but that is much faster than the vertical direction. Here, only the content of the register is shifted, but not all columns. In the horizontal direction, the ADC cycle time must be added to finally calculate the read-out time of one data point. The total read-out time in CCD standard mode is calculated by

$$t_{\text{Read-}n} = (SL \times t_{SL}) + (SR \times t_{SR}) \times (hb \times t_{ADC}), \quad (5.3)$$

where $t_{\text{Read-}n}$ is the total read-out time of a CCD in normal mode, SL is the number of existing vertical lines, t_{SL} is the time to shift one line vertically, SR is the number of horizontally existing register pixels, t_{SR} is the time to shift one register pixel horizontally, t_{ADC} is the time required for one ADC conversion, and hb is the number of data points (either single data or binned/grouped data—it is assumed that the storage of data in the computer's hardware creates no delay).

The background and noise of the pixels depend on the time between two read-outs and the sensor temperature, but only slightly on the size of a pixel, as shown in Table 5.2. The background at a given temperature increases linearly with time. Efficient cooling is therefore very useful. The gradient of the background is halved per 6–8-K temperature reduction. The noise of the background is the STD of the number of background electrons and depends on the pixel size. With increasing size, the capacity/noise ratio will improve. The same is true for binned pixels. For example, if a pixel has a 10-e^- background and 4-e^- noise, 10 binned pixels would have a 100-e^- background at only 12-e^- noise; the background noise is relatively smaller. A very-well-cooled CCD, selected for long-term integration, can carry a background and noise of between 1 and 2 e^- only after several minutes. When binning many pixels into a large cluster, called a superpixel, both the capacity and the background grow linearly with the number of pixels but not the noise. When

binning in vertical order, it is important to ensure that the total number of electrons in the cluster does not exceed the capacity of the register pixel. In the horizontal direction (the distribution of wavelength), the danger is less because in standard use the register is binned in the same fashion.

5.8.2.2 Applicability of CCDs for spectroscopy, image processing, and photography

For image processing, a dynamic range of 10^3 per pixel might be sufficient, and many small pixels might be preferable to process fine structures. For spectroscopy, even when hyperspectral, the requirements can be different. In any case, it is a good idea to review the pixel size, spatial resolution, and linear range of the complete system for the experimental needs. A good example might be the effectiveness of 12 or more megapixels at a 3/4-inch (18-mm) chip. At the common x/y ratio of 3:4, it has 4000 pixels at 15 mm, or a pixel size of $<4\text{ }\mu\text{m}$, and little capacity, realized by a very fast CMOS chip. Because a photo camera will not be cooled and because of the customer's expectations, fast read-out is important. All of these factors produce a dynamic range of, at best, 10 bits (1024:1). To improve matters, smoothing algorithms are directly applied—why not take larger pixels instantly? In an expensive 35-mm CCD camera, the same pixel number looks much more meaningful. The pixels will be $7\text{ }\mu\text{m}$ and produce a dynamic range of 12 bits, which will not require immediate smoothing. In general, a chip that is useful for photography will not automatically be useful for scientific imaging applications, and only in rare cases for spectroscopy.

5.8.3 Signal transfer and read-out

The read-out of diode arrays is rather simple: The diodes are read sequentially and are recharged at the same time. The result is transferred into a register, then to a pre-amplifier, to the ADC, and finally into the data storage.

The read-out of a 2D CCD is more complicated. A CCD consists of horizontal lines and vertical columns. The pixels fill the surface in both directions without gaps, but between the pixels run fine wires that are used to control the elements. Assigned specific voltages, each element can be addressed separately. The wires can also soak away excess electrons that occur at optical “overload.” Thus, the pixel crosstalk (“blooming”) is rare with scientific CCDs. The variation of the control voltage allows pixels to be combined into clusters to produce a “superpixel” or to transfer the charge from pixel to pixel (bucket chain principle). Imagine the process like this: During illumination, there are small walls between the pixels, which, in turn, can be considered as small pots (“wells”). The walls can be removed to get larger pots. The read-out happens according to a sequence of shifting processes. By varying the control voltages, each line of pixels or superpixels transfers its content to the next line, which requires two signals (clocks) per shift. At the start, pixel line 1 transfers to the register, which takes over the

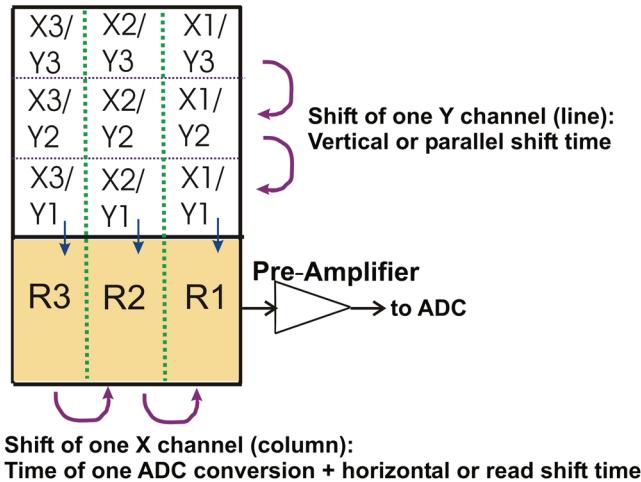


Figure 5.13 Shifting processes to read out a CCD.

content from line 2, which will take over from the next vertical neighbor, and so on. All lines are shifted towards the register, and each process is a vertical (parallel) shift. The time required for one shift is between 0.5–50 μ s per line, depending on the CCD model.

The CCD values in Table 5.2 are typical for the shift time, but large variations exist. CCDs trimmed for fast shifts and also those comprising large pixels can be obtained. They can have other drawbacks, such as lower efficiency. The first horizontal line in Fig. 5.13, called Y1, is connected to the read-out register R. The register, in principle, is another line of pixels that is covered from light. It often has larger elements than the active pixels, as shown, which increases the capacity and allows for the intermediate storage of the content of more than a single pixel. Many CCDs incorporate additional pixels (not shown) between R and line Y1, and probably also above Y_n, which are called dummy pixels. Their purpose is to buffer and synchronize the read-out. They are also found in the horizontal orientation before and after the register. In the calculation of transfer times, they need to be regarded as normal pixels. The horizontal movement (horizontal or read-out shift) from the register towards the ADC also needs two clock signals. However, because now only the register pixels are involved and not the active ones, that process is faster (10–500 ns per shift). Studying supplier data sheets might reveal that the horizontal shift time is included in the ADC conversion time. In the following calculation, it will always be mentioned.

Consider the read-out of the first line, Y1. The register is considered empty; no dummies are involved. On the clock, line 1 (pixel X1/Y1 through X_n/Y1) transfers its charge in parallel fashion to the register, and immediately after that, it takes over the content of line 2 from Y2. The time required is the vertical shift time. A time of 10 μ s is assumed for this example. Within that time, two processes take place: For clock 1, the line transfers the content to

the next line or to the register. For clock 2, the line considered takes over the charge from the neighbors above. All others follow. After all parallel shifts are completed, the horizontal shift, including the ADC conversion, begins. It shifts all horizontal register contents (but not that of the active pixel) by one channel in the direction R1. Considering an ADC conversion time of 10 μs (often also found as 100 kHz) and a horizontal shift time of 0.2 μs , one register/ADC process needs 10.2 μs . Equation (5.3) is applied [$t_{Read-n} = (SL \times t_{SL}) + (SR \times t_{SR}) \times (hb \times t_{ADC})$], and different cases of reading out a complete CCD with 512 pixels \times 512 pixels are reviewed. First, there is the “spectroscopy mode,” which means that all vertical pixels of one column are binned together, while the horizontals are separated. There is also “image mode,” which means that each pixel is read and stored separately.

To read and store one line takes [1 \times vertical shift time t_{SL}] + [$SR \times (\text{ADC time } t_{SR} + \text{horizontal shift time} \times t_{ADC})$]. That intermediate result, multiplied by the number of lines SL , produces the whole read-out time. A CCD with the parameters considered so far, and 512 pixels \times 512 pixels, needs $512 \times 10 \mu\text{s}$, or 5.12 ms, for all vertical shifts, plus $10.2 \mu\text{s} \times 512$, or 5.2224 ms, for all horizontal shifts including AD conversion (assuming that storage costs no extra time). That results in 10.3434 ms for one complete read-out in “spectroscopy mode,” which also represents the shortest repetition rate of spectra. In the fully separating “imaging mode,” all vertical shifts are also found at 5.12 ms, but the horizontals are now ($10.2 \mu\text{s} \times 512 \times 512$), equaling 2,673.8688 ms for all horizontal shifts and conversions and resulting in a total of 2,689.1688 ms for one full image. In short, the time required for the parallel shifts in spectroscopy mode is, in the example, ~50% of the total read-out time; however, in imaging mode, it is small (0.2%). There are ways to vary the time required by special binning and special configuration modes; see Section 5.8.8. High-performance scientific CCD cameras allow access to and variation of the vertical/parallel shift time. However, reducing the shift times below the minimum proposed by the manufacturer can lead to inefficient transfer and bad linearity. It is a good idea to test and decide upon requirements about what parameters are most important: time, spatial or spectral resolution, or quality of data.

Reading out a diode array is simple: ADC conversion time \times number of pixels or superpixels. Illumination preparation considers several factors. The experiment often delivers light as constant flux. Depending on the cooling of the sensor and stray light, it might be beneficial to keep the detector “virginal.” The process to achieve that is called “cleaning,” “keeping clean,” or similar. The function erases the detector without the storage of data. Depending on the user’s comfort with CCD and camera, there are three methods. The simplest method shifts all vertical contents over and erases the register after each complete set of shifts. A set consists of the normal vertical and a single horizontal shift time. Other CCD models allow for much quicker erasure by an automatic change to a faster shift time upon starting the erasure

process. Finally, some CCDs feature a “flash clock” that, if triggered, erases the complete content within microseconds. Of course, if the CCD can accommodate this feature, the camera’s hardware and software must be able to control the different special functions. (Section 5.8.6 discusses specialization further.)

5.8.3.1 Combining the read-out in imaging mode and the display in spectroscopy mode

With vertical binning (spectroscopy mode), the capacity of the register pixels can quickly be reached, after which nonlinear results would follow. On the other hand, it might be useful to receive large numbers of counts (photons) for low noise, as with quantitative data reduction. The solution is to read the pixels separately or in small vertical groups, and, after ADC processing, add them again and display them in a spectral fashion. That process is called software binning. The read-out is in imaging mode and requires rather-long read-out times.

5.8.4 CCD architectures

Three different kinds of CCD architecture are on the market. Each one offers advantages and disadvantages for spectroscopic applications; the easiest to describe is the full-frame architecture, displayed in the center of Fig. 5.14. In spectroscopy, it is the most popular version.

The principles of charge transfer fully apply, as described in Section 5.8.3. Almost the full area is available for exposure—it has the highest area efficiency. During read-out, additional light is allowed to the surface only if all vertical information (pixel columns) is binned into a single data point

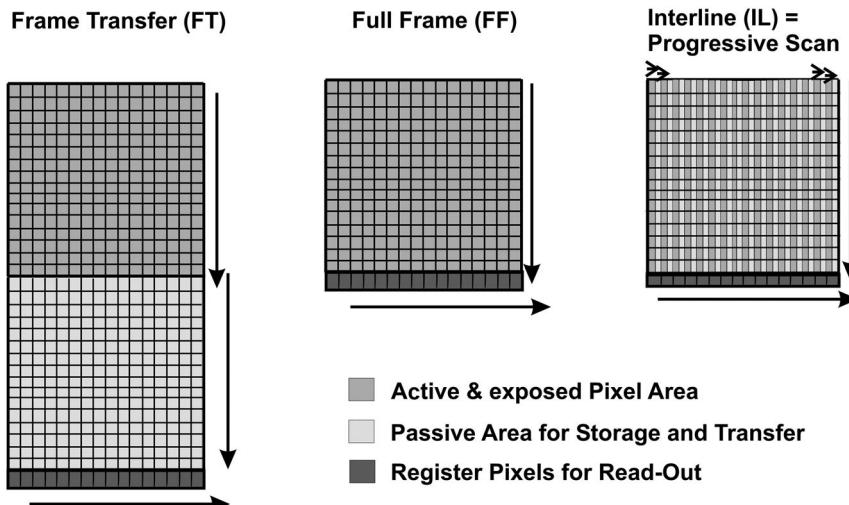


Figure 5.14 The principal CCD architecture.

(spectroscopy mode) either by hardware or by software. That is fine so long as the experiment light carries no 2D information; however, the full height is often not illuminated, and the unused areas are swapped. That scenario requires at least two regions of interest (ROI) and at least two read-out processes, even if only one read-out is digitized and stored. That primitive allocation already forbids exposure during read-out to avoid crosstalk, which can produce false results. Multi-track spectroscopy strictly requires read-out without any additional illumination. Many full-frame CCD models comprise two registers, one on the upper end and one on the lower end of the active area. One can be optimized for high capacity and speed, while the other is better suited for a low noise and detection limit. The change between the two is addressed by the camera program. Even models with four registers (one at each side) are available, the fastest of which allow the read-out of all four in parallel. Those models have been developed for high-speed read-out applications.

The left image in Fig. 5.14 shows the frame-transfer principle. It consists of a rectangular CCD, one half of which is subject to illumination, and the other half is protected from light and used as a passive intermediate store. At the end of the exposure time, the content of the active (dark grey) part is shifted into the storage region (light grey). While the active part might be exposed again, the passive one is read out; that shortens the transfer time and improves the ratio between exposure and read-out (duty cycle). Advantage number two is the option to make two exposures quickly, with a transfer in between, and read out the whole content after the second exposure. Advantage three is the application of kinetic measurements with many exposures within a very short time. Both applications are described in Section 5.8.10. Frame-transfer CCDs are available with optimized vertical shift times, reaching below 1 μs per line. The read-out process works according to Eq. (5.3); however, one must keep in mind the additional lines of the covered storage area, which normally double the number of line shifts. It is also important that during read-out only the lines of the storage area are shifted. The lines of the active area are kept until they are transferred into storage, which requires special options in the camera's hardware and software. Whether the handover from the active part to the passive area requires blocking of the light flux depends on the time required and level of "smear" that is acceptable.

The third architecture is the interline or progressive-scan CCD, shown at the right of Fig. 5.14. In the horizontal orientation, there is alternately one active column and one passive (and covered) column. On a clock, all charges from the active pixels are moved to the adjacent passive pixels at once. The read-out process is then done. The typical shift time for the one-step process is below 1 μs , which is so fast that in most cases it makes no difference if the experiment light remains on. While the old information is read out, the next exposure takes place, which makes the technique very time efficient. Drawbacks come from three parameters: Only half the surface is active. For a partial compensation, some interline CCDs have integrated microlenses

that guide a part of the light that would hit a passive pixel to an active one. This feature makes the pixels appear asymmetric. Furthermore, this architecture is only available with front illumination, which is not efficient in the UV. Finally, all interline CCDs comprise rather small pixel sizes, making them useful in spectroscopy only for a very limited number of applications. They are most often found in image processing with fast repetition rates and reduced dynamic ranges. The read-out process also follows Eq. (5.3). For each read-out sequence, the quick active-to-passive transfer needs to be added once.

5.8.5 CCD and array efficiency

5.8.5.1 Front-illuminated CCDs

Silicon is one of the most efficient materials in the visible wavelength range. Its quantum efficiency (QE, yield) spans up to almost 90%. Its structure and endowment in the light-sensitive diode affect the behavior of the efficiency and allow optimization to certain wavelength ranges. The structures of a CCD are grown on quartz substrate. The structures are protected from the environment by a transmitting layer that absorbs in the UV. Fine wires run between the structures; the wires reduce the illuminated area per pixel. Figure 5.15 shows typical curves of front-illuminated CCDs. Note that the real curves might differ from model to model.

The basic shape of the efficiency curve is similar for CCDs with frame-transfer and full-frame architecture. The standard version is marked dark blue. If a scintillator is added, the curve extends into the UV (blue dashed). A modified version is the “deep depletion” with a different endowment and deeper structures. The efficiency is then shifted towards longer wavelengths (brown and dotted brown). If the safety layer is removed (the element must always be kept under vacuum), the overall efficiency increases, and the UV range becomes effective (green), which is called an “open electrode CCD.” CCDs with interline technology greatly differ, as shown in the red curves. Their additional structures and optional microlenses create interference. They are available with UV scintillators, as well.

5.8.5.2 Rear-side-illuminated CCDs

If a CCD is reversed so that light shines on the side of the substrate, the light must not pass either the absorbing structures or the UV-absorbing layer. The efficiency curve now shows its “true face.” Rear-exposed CCDs have thinned substrates to reduce the absorption; unfortunately, that thinning will create interference, which becomes noticeable between ~ 700 – 1000 nm. A variety of rear-side CCDs are available, comprising different efficiency behaviors.

As shown in Fig. 5.16, CCDs with a normal structure and endowment will be close to the turquoise standard curve. A scintillator increases the UV efficiency reasonably well (turquoise dashed). The “UV-enhanced” curve

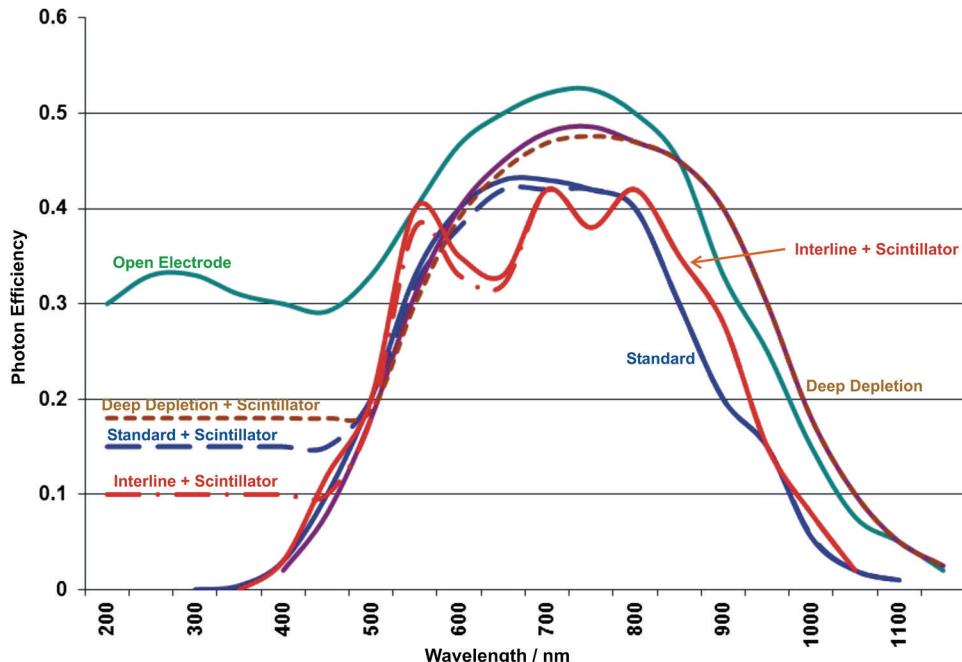


Figure 5.15 Typical efficiency curves of different types of front-illuminated CCDs.

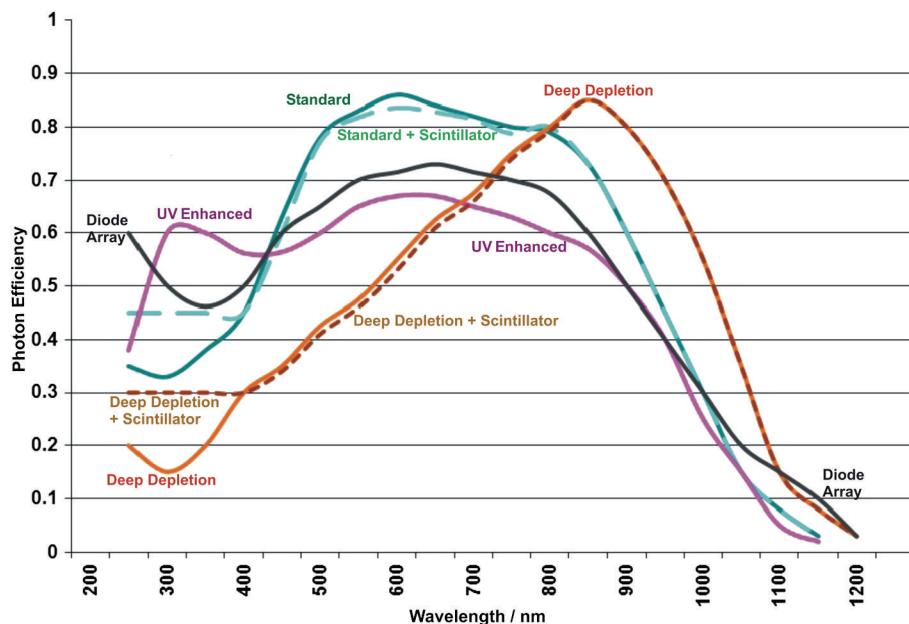


Figure 5.16 Typical efficiency curves of rear-side illuminated CCDs.

(violet) shows how varying the endowment can also change the wavelength-dependent behavior. Here, the UV is strongly improved without a scintillator at the expense of the middle and upper wavelengths. For NIR applications (yellow curve), deep-depletion versions might be preferable because they have a thicker substrate (which absorbs more strongly in the UV) and the structures are etched deeper. The light will penetrate deeper (hence the name). The whole curve is shifted towards the NIR, and, importantly, the interference is strongly reduced. Scintillators are also applicable (brown dotted). The substrate of a rear-side-illuminated CCD creates optical interference (see Fig. 5.18).

The standard efficiency curve of Si diode arrays in Fig. 5.17 is the black curve in Fig. 5.16. The NIR part is the grey dashed line in Fig. 5.17. The other three are typical InGaAs curves. The wavelength range of UV silicon reaches from below 200 nm to 1100 nm; the range above 750 nm is shown dashed for reference. The range of standard InGaAs is 700–1650 nm (IGA 1.7, solid black curve). Like the single-element sensors, there are also IR versions with upper limits at 2200 nm (IGA 2.2, dotted) or even 2700 nm (IGA2.7, dashed black). The arrays and CCDs beyond the Si range are all hybrids. The substrate is a Si chip, upon which the active material is grown. The created electrons are transferred from the active material to the Si substrate for further storage and transfer. Many different hybrids are offered. Historically, dual Si-diode arrays have been marketed to measure two separate spectra, one on top of the other. However, they disappeared due to the higher flexibility of the 2D CCD. Diode arrays are read in a single sequence after the start of the read-out process. With binning (in array technology, the word “grouping” is preferred),

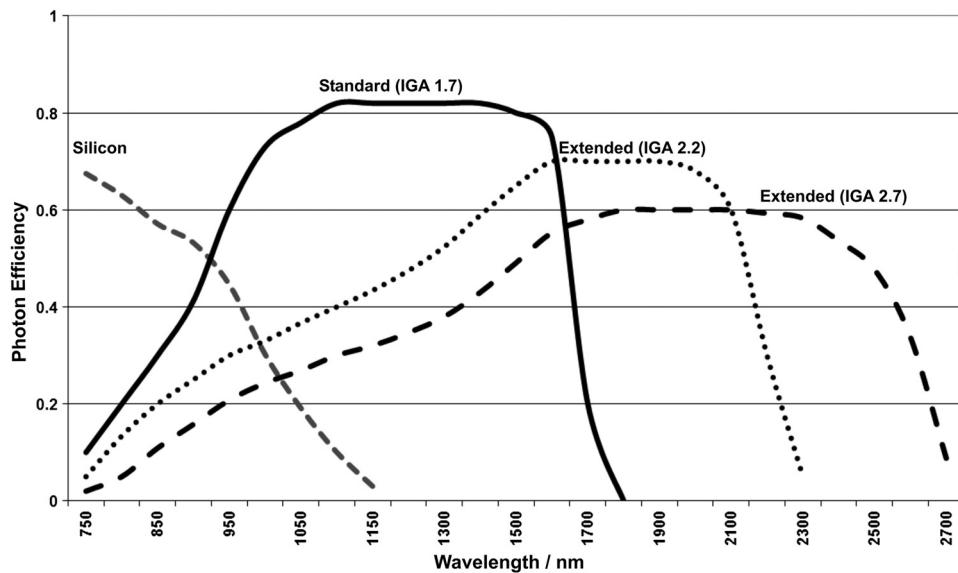


Figure 5.17 Three representative indium-gallium-arsenide efficiency curves.

it is possible to combine a number of diodes into a single ADC, which would reduce the background and quicken the read-out.

5.8.5.3 Interference of rear-side-illuminated CCDs: Etaloning

Rear-side-illuminated CCDs produce interference. The light passes through the thin substrate to reach the active layer; thus, two changes of refractive index are passed, and reflections occur. The interfaces are sharp, and the substrate is thin. The effect is called etaloning.

Although the three types with front illumination shown in Fig. 5.18—standard front, front deep depletion, and front open electrode—do not suffer from interference, the effects are unavoidable with rear illumination. The systems most affected are standard rear-side-illuminated models. The modulation can occur between 750–950 nm and will be strongest in the range of 800–900 nm. The period in the graph is shown to be 20 nm, but in reality it can be even shorter than 5 nm. The amplitude can reach up to 30%. The parameters are the substrate thickness, the sharpness of the interface between substrate and element, the endowment, the angles of the arriving light, and the spectral content of the light (a laser will create hardly any interference). The deep-depletion version creates far less interference. Under the same experimental circumstances, the period will be remarkably longer and the amplitude will, in most cases, stay below 10%. If the experiment does not tolerate any interference, front-illuminated CCDs must be applied.

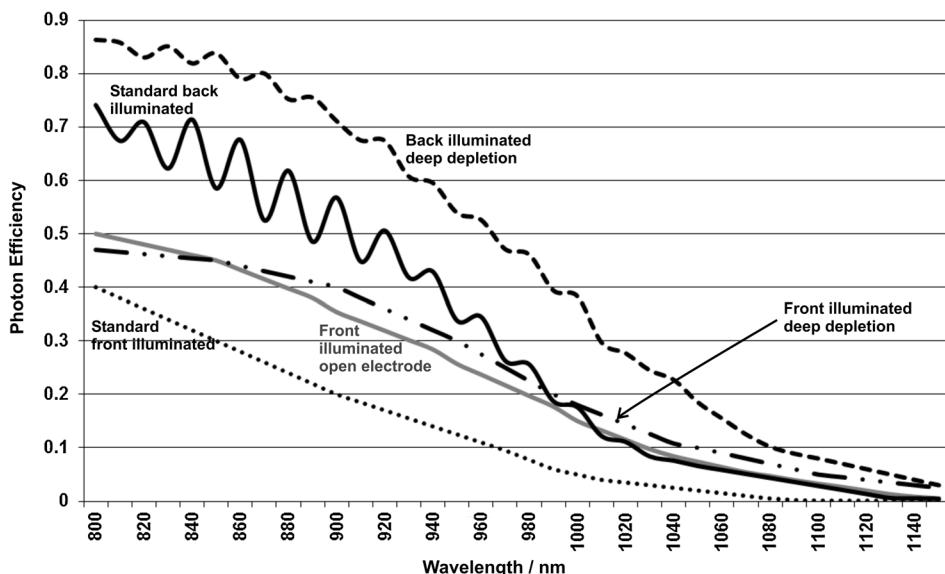


Figure 5.18 Typical CCD behavior in the NIR.

5.8.6 Time control: synchronization, shutter, and gating

All scientific CCD cameras can be triggered by the experiment. The read-out process will start after the trigger and after a pre-programmed delay time. It might also be possible to perform a triggered cleaning process and store the next set of data according to the read-out definition. If another trigger arrives during the exposure time and read-out, it will have no effect. Very good camera systems accept an additional external synchronization signal besides the trigger, which provides several advantages, such as starting the read-out only after both independent signals have arrived or if the trigger starts the whole experiment, and each sync signal controls one read-out. For example, suppose a camera has been started by a constantly high main trigger after the first sync signal. After a read-out, the camera goes into “keep-clean” mode, which wipes away all charges. After the sync, the cleaning will stop (wherever it is), and the integration time begins; following that, the camera will read-out by the pre-defined mode. If, after the read-out, the main trigger is still high, it will clean again until the next sync arrives, and so on. The negative slope of the main trigger will stop the whole sequence.

In the other direction, a scientific camera can synchronize the experiment, control it, and trigger it. That ability also requires two independent signals: one might be provided in front of or after the read-out, and the other at a defined point inside the read-out process (thus allowing a laser to be triggered in the middle of exposure). Trigger-in and trigger-out are also combined. The system can, when conducted with a read-out, send out a trigger and go to keep-clean mode. After receiving the external answer, the integration and read-out will start again. Another advantage of using an external trigger combined with an external sync signal is discovered if one does not know how long an experiment could last or need to be completed. Intelligent systems will allow one to define that data will be taken so long as the main trigger is high and each sync starts a measurement. Upon a falling trigger, the whole experiment finishes, and the number of sequences is made available in a register. Vertical separation of optical signals requires shifting the charge in a sequential manner. Therefore, no additional photons must reach the detector during read-out. Some questions lead to the method of choice: Can the light source be synchronized (stroboscope)? Can the light source be pulsed and synchronize the detector in return? Is the detector itself able to interrupt detection (MCP image intensifier or other methods)? Can one or more shutters be placed in the light travel, eventually near the spectrometer’s entrance slit, and be controlled by the detector electronics?

5.8.6.1 Shutter control

For all kinds of arrays and CCD detectors, the cycle time is defined by the exposure time plus the shift and read-out times. The detector always “views” light. In experiments providing constant illumination, that factor can lead to problems, especially when reading out in 2D mode. Separating the exposure

and read-out can be done by several methods, including provisions to open and close the light path using an electromechanical shutter. Depending on size, a shutter needs between 5–50 ms to open or close. Thus, in spectroscopy it is very useful to mount the shutter on a place where the shutter's lapse time is the shortest possible. In most cases, that place is the entrance slit of the spectrograph; the maximum aperture is often limited to 3 mm × 3 mm. Shutters are generally constructed as an iris aperture with an open aperture of 18 mm. If it needs 20 ms for one opening or closing process, it will “wipe” over the height of 3 mm within 3.3 ms. That number will be the blur in time that needs to be accepted. At least it will be reproducible. The repetition rate of a shutter is limited. More than 2–4 cycles per second should not be expected from standard versions; otherwise, thermal effects (jamming) or even long-term damage can occur. A positive side effect of a shutter is saving the spectrometer and the detector from unwanted illumination as long as the system is turned off or inactive.

5.8.6.2 Microchannel-plate image intensifiers

For shorter exposure times and/or higher cycle frequencies, the use of gated image intensifiers might be a solution to separate exposure from read-out. Section 5.4 and Fig. 5.5 describe the MCP. It can be used as a very fast electronic shutter in front of a CCD. The system performance will change completely, and the efficiency curve will be that of the MCP photocathode instead of that of the CCD. This change can be either advantageous or detrimental. The shutter times performed by an MCP/CCD system can be as fast as 2 ns or even shorter. The repetition rates of the MCP gate reach well into the kHz range, opening lots of new applications for synchronized and pulsed experiments, which will be discussed next.

5.8.7 Current formats of area detectors

Diode arrays are always line-shaped. For the applications discussed here, the single elements can be a quadratic shape of 25 μm or larger. They can also be rectangular, which offers advantages for spectroscopy. Widths of 25 or 50 μm, and heights of 2.5 mm are often used. The number of elements spans between 64 and 2048. Like CCDs, they can work in the visible, NIR, or be expanded into the UV. Extension into the deep UV also works by adding a scintillator (UV phosphor). Light coupling via fiber optics is available and proven. Arrays with integrated fiber optic brushes allow adding a scintillator, connecting another fiber optic brush, or coupling the light therein. Fiber optic coupling changes the format or combines with an MCP. After starting the read-out, diode arrays are read out in one sequence (self-scanned). Elements can be combined into common data points (binning or grouping), but that does not deliver as big an improvement in SNR that a CCD would. Regardless, the background is slightly reduced, and the read-out speed increases linearly. The background increases with increasing IR efficiency.

CCD sensors are marketed in a large number of versions. In most cases, the pixels are of quadratic shape and reach from $<5\text{ }\mu\text{m}$ to $>50\text{ }\mu\text{m}$. The chip formats cover the duplication of diode arrays (e.g., 1300 pixels of $20\text{ }\mu\text{m}$ per line, and 100–400 lines high) that are ideal for standard spectroscopy. They can even fit spectrographs without image correction. Rectangular or squared formats, besides the designated imaging applications, are useful for multi-stripe spectroscopy and hyperspectral applications (see Section 4.5), including kinetic spectroscopy. CCD detectors, if used at the appropriate sensor temperature, are very quiet systems. The noise is almost completely created in the chain register/pre-amplifier/ADC, but not at the CCD chip itself. Consequently, it is a good idea to program the minimum number of conversion processes for a measurement, while saving the register and ADC from overload. Additional reasons include the time advantage and reduced memory load. As an example, consider a read-out noise of 5 e^- . A single pixel can add 1 e^- to the noise of the read-out chain. Binning 4 pixels \times 4 pixels will probably increase that number to 1.02 e^- , and binning 100 pixels can lead to 1.2 e^- . It should be noted that if good cooling is provided, the noise from the CCD itself can be neglected; it will add, at most, no extra background (see the CCD parameter in Table 5.2). That behavior greatly distinguishes the CCD from single-element detectors: while increasing the active area (by binning), the CCD detector noise will almost stay constant. In fact, this is one of the undisputable strengths of CCD technology!

5.8.8 Read-out techniques: Multi-spectra spectroscopy, binning, and virtual CCD partition

To acquire several different spectra in parallel, they must be guided and read separately. Thus, at the spectrograph's entrance, they must be separated vertically, which can be achieved by a faceplate with different holes, by fiber optics, or by mirrors. The spectrometer must be able to reproduce the spatial separation in the output. From the detector system, it is expected that it also can process the spectral information free from crosstalk and interference. The spectrometer can provide intrinsic image capability or comprise internal or external imaging correction. The following example assumes perfect internal correction. The CCD comprises 1024 pixels \times 1024 pixels with a 24.576-mm height and a 24.576-mm width. The entrance slit provides a height of 15 mm, and four parallel light signals require 12 mm of that height. These signals originate from four fiber tips that illuminate the entrance slit with the correct cone angles, which, in turn, is provided by additional optics between the fibers and slit. Figure 5.19 presents a cross-section of the system and its light travel.

Four superposed fiber guides illuminate the entrance slit. Two of the beams are displayed and marked in blue and red. That way, they enter the spectrometer vertically separated before exiting the optics and hitting the CCD detector (also vertically separated). Whether a vertical inversion happens, as shown in the graph, is just a matter of the number of optical

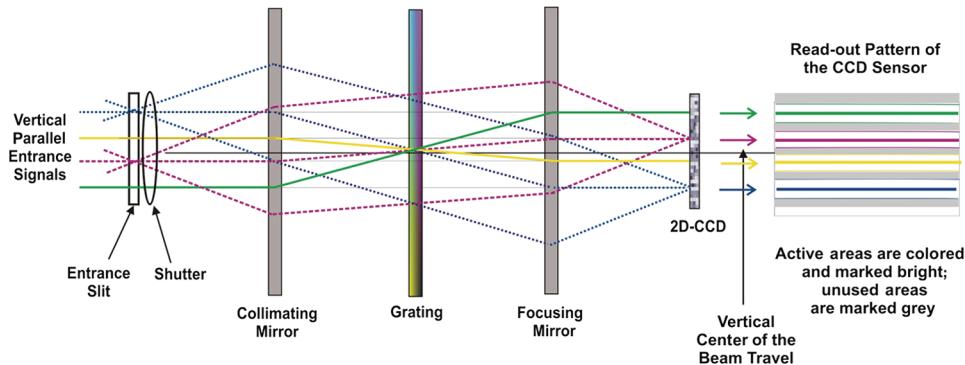


Figure 5.19 Example of the beam travel in a multi-spectra experiment.

components and reflections. For simplification, all mirrors and the grating are illustrated in transmission.

Detailed description

The vertical distance between the two inner optical signals is ± 1.5 mm with respect to the center, while the outer signals arrive at ± 6 mm. No signal hits the very center; that would occur if an odd number of beams enter. For the upper (blue) and the third (red) beam, the vertical spread is shown, while for the second and fourth beam only the center travel is drawn. Consider the blue signal's travel: It enters at position +6 mm and hits the collimating mirror at the same height. From there it is reflected in a collimated fashion. Due to the main curvature of the collimator, it then travels downward and hits the grating in the center position (like all other beams). Under the best conditions, the spread of each beam fills the grating surface. The focusing mirror is hit at -6 mm. Because the focusing mirror is basically symmetric to the collimator, the light will travel farther at a parallel height to the detector, arriving in a refocused shape at -6 mm. All other beams behave comparably. They pass the grating in the middle and hit the detector at the inversed height of the starting position at the entrance. Along the way, they form parallelograms; ideally, they would reproduce the shape of the entrance spot in the plane of the detector. The ideal reproduction of, for instance, a 100- μm -high spot, would be the same size in the spectrometer output. As already stated, that will not always be the case. The higher the optical quality Q_i of the spectrometer (see Section 2.6.6), the sharper and more precise the reproduction. At the CCD detector, the height of 24.6 mm is divided into nine horizontal stripes of different heights. The four stripes that receive the optical signals are marked and have vertical reserve in both directions to tolerate the aberrations of the spectrometers. Because the aberrations increase with the distance from the center, the upper and lower bands (stripes) are designed slightly wider

than the two central bands. The four stripes are separated by passive bands (marked grey in Fig. 5.19) that are not AD converted during the read-out.

The read-out pattern of the CCD follows the illumination. According to the position and height of the signal stripes, all vertical pixels are binned into superpixels, resulting in four separated spectra. The content of the passive pixel bands is shifted through and erased. Figure 5.20 illustrates the pattern.

The two inner spectra cover less than 40 pixels (<1 mm) in height. The two outer spectra are allowed to use slightly more vertical spread: 50 pixels (1.2 mm). Between all used stripes are 60 unused pixels (grey, reserve), extended to 80 in the center. The ranges above and below ± 6 mm from the center are unused. In total, there are nine bands (ROIs). The register is found at the lower boundary of the CCD.

A condition for linear measurements is that the sum of all charge carriers in each measured and “binned” band must not exceed the capacity of the register and that of the ADC. The parallel shift time is 8 μ s, the ADC needs 9.8 μ s, and the register shift time is 0.2 μ s. In order to avoid data overlap, the system contains a shutter right after the entrance slit, which closes during the read-out.

According to Eq. (5.3), the timing is as follows: The lower grey band, with a height of 232 pixels, is shifted within 1.856 ms, and the register is erased in 0.2048 ms. The blue band needs 0.64 ms to enter the register and 10.24 ms for horizontal shift and conversion. The second reserve, with 60 pixels, is shifted within 0.48 ms into the register and is also erased within 0.2048 ms. The yellow track has 60 vertical pixels, and it needs 0.48 ms to reach the register and 10.24 ms for conversion and storage. In the center is a reserve of 80 pixels that requires 0.64 ms for the shift, plus 0.2048 ms erasure time. The lower half of the CCD is then processed, and the sequence repeats symmetrically for the

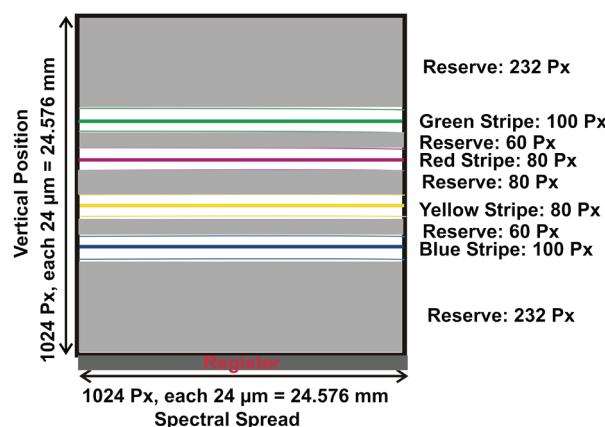


Figure 5.20 Read-out pattern of a square-shaped CCD for multi-track spectroscopy.

upper part. In short, the total time for all shifts and ADC conversions in the example is 50.176 ms. The share of 40.96 ms is used for the conversion of the four spectra, 8.192 ms are required for the vertical shifts, and 1.024 ms are used for the horizontal erasure of the register.

After that time, the electronics would be ready for the next exposure, but the shutter is not fast enough. More than four cycles per second (CPS) are not reasonable with an electromechanical shutter. If the experiment requires the technically possible ~ 20 measurements/s, another kind of discrimination between the exposure and read-out needs to be chosen (see Section 5.8.6).

Supposing the question of repetition rate and shutter time is solved, as would be the case with a stroboscopic light source, what can be done, to speed-up the read-out of the CCD? The first answer is to use a faster ADC. As Table 5.2 demonstrates, a CCD with a pixel size of $24 \mu\text{m} \times 24 \mu\text{m}$ provides a dynamic linearity of 16 bits (65,536 units) at a conversion time of 10 μs . Consider an ADC time of 1 μs (including the register shift) and accept that the dynamic linearity will drop to 22,000 units. The scenario will not pose a problem because all vertical pixels are binned in a stripe. Therefore, more photons can be combined per read-out to reach the required SNR. The total vertical shift time will stay at $1024 \times 8 \mu\text{s} = 8.192 \text{ ms}$; adding five erasure processes of $5 \times 1024 \times 0.2 \mu\text{s} = 1.024 \text{ ms}$ and then the new ADC conversion time for four spectra, $4 \times 1024 \times 1 \mu\text{s} = 4.096 \text{ ms}$, sums to 13.312 ms per cycle, compared to 50.176 ms at a 10- μs conversion time. At the CCD, 75 measurements/s would be possible. If that still is not fast enough, the next answer is to use virtual programming.

5.8.8.1 Virtual CCD programming

Particularly highly developed scientific CCD cameras allow one to split the CCD sensor into separate, independent regions: virtual CCDs. In the example presented in this section, the CCD would be divided into nine virtual CCDs according to the nine vertical bands. The five bands marked in grey, which receive no experimental light, become one superpixel each. The vertical and the horizontal shifts of those unused superpixels will then need an estimated time of $8 \mu\text{s} + 0.2 \mu\text{s}$ each, for a total of 41 μs . Upon calculation, it will be necessary to set up the experiment and run the read-out to measure the timing precisely because every CCD will have a slightly different personal-time behavior. The stripes with useful signals will remain unchanged in terms of shifting and conversion to maintain good precision. At an ADC rate of 1 μs , the numbers would be $360 \times 8 \mu\text{s} = 1.440 \text{ ms}$ for the vertical shifts, plus $4 \times 1024 \mu\text{s} = 4.096 \text{ ms}$ for the read-out of the data, plus 41 μs for the unused superpixels, which produces an estimated total time of 5.577 ms per cycle and allows 179 measurements/s. If that level is still too slow, it is possible to pursue horizontal binning at the expense of spectral resolution. Optionally, the superpixel fast-shift method can also be partly utilized with the data channels. Tests will indicate whether the linearity is sufficient.

Even with the methods presented so far, there is the danger of losing charge carriers during the parallel shifts, damaging the linearity and the repeatability. Shifting too fast or too many charge carriers at once can lead to residual information in the passed pixels, which changes the background and transferred carriers in a critical manner. However, many kinds of modification and optimization are possible with virtual programming. The definition of virtual CCDs offers even more access to different parameters, such as special exposure times for certain regions, virtual shifting of x - y blocks instead of shifting whole lines, and more. CCDs on the market can comprise two or even four registers. Virtual programming now allows one to move part of the sensor to the different registers in parallel and read out two or four ADCs. Because of the many varying parameters and options, it is not possible to develop an equation for the read-out times of modes like that. The virtual mode also is attractive because it allows one to shift and read one part of the sensor while keeping the other in place. Of course, regions “hidden behind other regions” can be moved separately only after those in front have been shifted away. That behavior permits very complex exposure and read-out patterns for CCDs with more than one register. Finally, if fast and direct access is preferred over data quality, there is a competitive detector: the CMOS (see Section 5.9.1).

5.8.9 CCDs and array systems with image intensification

Weak light signals, which might be available for short times only, are hard to integrate over time, but they can probably be acquired. Several methods for amplification have been developed in the past; currently, only two are routinely used: die electronic multiplication of charge carriers at the CCD chip (on-chip multiplication), and the application of MCP image intensifiers in front of the CCD or array. It is also possible to combine the two.

5.8.9.1 CCDs with on-chip multiplication or electron multiplication (EMCCD)

An (otherwise normal) frame-transfer or full-frame CCD with front or rear illumination is equipped with one or two altered registers. The modified register then can transfer the charge in both directions. The “normal” direction shifts the electrons towards the pre-amplifier and ADC. In the “extra” direction, a number of additional elements are added between the register and pre-amplifier/ADC. The additional elements include an adjustable “acceleration voltage” that effects a small but reproducible electronic multiplication (EM) factor between 1.00 and about 1.03 each. If, for example, 256 of the EM amplifiers work in a row, an adjustable and reproducible gain of 1–1600-fold is provided. This behavior implies an advantage in creating a gain on the cooled chip without the pre-amplifier noise. Thus, the read-out noise increases very little while the useful signal increases by a large factor. A typical value for the normal read-out noise of a 1- μ s ADC time and a pixel capacity of 200,000 e[−] is around 10 e[−]. At an EM

amplification factor of 600, the noise increases to $\sim 40 \text{ e}^-$, providing an SNR profit of factor 15 in the end. Care must be taken to avoid saturation effects that can appear rather quickly. Furthermore, at very weak signal levels, very high gain makes little sense because of the strong natural variation in the photon statistics. In practice, a gain factor of ~ 700 makes a reasonable maximum. Because the data are shifted out of the register into different directions, the images or spectra will flip in the vertical axis when changing from normal to EM mode. Apart from that, an EMCCD behaves like a standard CCD.

5.8.9.2 CCDs with an additional microchannel-plate image intensifier (MCP-CCD)

The application of an MCP creates completely new, additional measurement modes and changes the principle of CCD detection. (The MCP is described in detail in Section 5.4.) The combination MCP-CCD, in most cases, is accomplished by optical fiber brushes. As described in Section 5.7.4 (Fig. 5.12), the fiber optic translation may accomplish a modified size, or shape, of image. A coupler performing this is called a “fiber taper.” Taper ratios between 2:1 and 1:2 are common. Relevant MCPs mostly come with a channel cross-section of $\sim 6 \mu\text{m}$; adding the thickness of the intermediate fiber walls, the median distance is between $7\text{--}8 \mu\text{m}$. A phosphor follows the output of the plate to convert the electrons into photons again and send them into another fiber brush, integrated into the phosphor. The fibers can again be a taper that ends on the opposite side in a CCD with pixels of, e.g., $13.5 \mu\text{m} \times 13.5 \mu\text{m}$. That setup provides a median of six fibers at one pixel, which gives a good averaging effect. The larger the pixel is, the better the averaging. A CCD with $13.5 \mu\text{m}$ pixels and 1024×1024 channels has a diagonal dimension of 19.5 mm. If it is connected to an 18-mm MCP and a taper translation of 1:1.1, the CCD is fully used. If a straight fiber brush is used, the extreme corners of the CCD will see no light. Alternatively, a 25-mm MCP with a taper ratio of 1:0.78 can be used to illuminate the 18-mm CCD. The decision about the 25-mm version can be based on the larger spectral interval at its entrance. It is generally better to have an image reduction between the MCP’s phosphor and the CCD because of the phosphor crosstalk, which creates slight blur effects that finally reduce the spatial resolution. The phosphor’s emission color should be as close as possible to the optimal wavelength of the CCD, and the decay time needs to be aligned with the experimental needs. The whole system’s efficiency will appear like a photomultiplier because it includes very similar cathodes. Besides the amplification of optical signals, the option of fast switching, called “gating,” is a major strength of MCP-combined systems. The number of electrons at the channel plate output that reach the CCD after conversion into photons is strongly reduced by the spherical emission characteristics of the phosphors. On the other hand, because several MCP channels illuminate one pixel, the gain in most cases is still enough to saturate the CCD pixel.

Typical MCP-CCD applications are pulsed-laser-driven experiments wherein the gate is synchronized with the laser. During the read-out, the gate is kept closed, and the laser is probably even stopped. To improve the SNR, the number of laser pulses per read can be pre-defined. The “on-chip” accumulation can be adjusted so that an optimal SNR is reached. Camera software will also provide the option to shift the time position of the gate relative to the trigger during acquisition. That enables the automatic recording of time and spectra, called 3D plots. The system linearity of a MCP-CCD setup depends on a large number of parameters. The transfer chain of a pure CCD system is light > conversion into electrons > shift to the register and amplifier > conversion of electrons into voltage > ADC conversion > storage. In a combined system the chain is light > conversion from photons to e^- > amplification inside an MCP > conversion from e^- to photons inside the phosphor > addition of some spatial distribution > the CCD chain described earlier. It is predictable that both the linearity and reproducibility in an MCP-CCD will be somewhat worse compared to the bare CCD. The background of a CCD is practically negligible, provided proper cooling for the time between read-outs. The MCP then creates thermal noise and shot noise at the photocathode, like a PMT does. Those noise signals are the same as those coming from the experiment; thus, the background increases linearly with time in an MCP-CCD system. Even though it will be reproducible and can be subtracted, its STD will increase. Even with the gate “off” the MCP will, due to the internal gain, create electrons in the channels and amplify them; however, they are remarkably fewer in comparison with the thermal e^- from the cathode, and, because they emerge from “underway” in the channel, they produce a smaller amplitude than the “real” signals they contribute. These facts prevent the reasonable use of MCP-CCD for true long-term measurements (~1 minute is the limit). In that respect, CCD and EM-CCD are dominant.

Reviewing the different parameters suggests that one could set up a modular system coupling between the MCP and CCD by classical optics. It would have stronger losses in intensity and optical aberrations compared to a fiber-coupled system. The rewards are that the image intensifier can be removed easily and at any time, and both the CCD and the phosphor can be run independently at their optimal temperature. The CCD must be an EM version to have all kinds of combinations available. Modular systems of that kind offer the ultimate flexibility and the best specifications under all circumstances at the expense of “ease-of-use.”

5.8.10 Data acquisition in the ms– μ s time frame

5.8.10.1 Kinetic measurements

The expression “kinetics” describes experiments with important parametric changes in the spectrum or image in the micro- to millisecond range. A primitive kinetics example is the decay of a phosphor. For spectroscopy

with a diode array providing 512 channels and an ADC with a $0.5\text{-}\mu\text{s}$ conversion time, a repetition rate (rep rate) of $256\text{ }\mu\text{s}$ without further binning is standard. For many kinetic experiments, that scenario will be sufficient. CCDs offer many options in timing. Some sacrifice performance, such as a reduced dynamic range. This section reviews some solutions for a CCD with 512 spectral channels. Frame-transfer CCDs, optimized for fast parallel shifts, are available, combined with a reasonable pixel size and dynamic range. A popular version has 512 horizontal and, because of the FT architecture, 1024 vertical channels. The pixel size is $19\text{ }\mu\text{m} \times 19\text{ }\mu\text{m}$, and the capacity per pixel is $>4 \times 10^5\text{ e}^-$, all at a nominal parallel shift time of $1.6\text{ }\mu\text{s}$. When choosing a standard spectroscopy read-out mode and binning all vertical 512 channels, the total transfer of all parallel lines into the passive part will be $819.2\text{ }\mu\text{s}$. Shifting from there into the register again requires $819.2\text{ }\mu\text{s}$. Together, the horizontal shifting and read-out at $0.5\text{ }\mu\text{s}/\text{data point}$ need $256\text{ }\mu\text{s}$. Thus, one complete cycle is a minimum of 1.8944 ms . That value is also the maximum rep rate ($527/\text{s}$) at full resolution. If some loss in dynamic linearity is acceptable, utilizing the virtual CCD mode and shifting all 512 lines as fast as technically possible can reduce all parallel shifts to $32\text{ }\mu\text{s}$, producing a rep rate of $\sim 320\text{ }\mu\text{s}$. So far, the FT-CCD is still the standard device. To reach faster time resolutions, the CCD can be modified. The majority of lines of the normally active surface can be covered and used for additional intermediate storage. To illustrate that principle, consider a situation with only 16 active vertical channels left. They must be located opposite the register (called the top area), while the register is the bottom area, as shown in Fig. 5.21.

The illuminated area is strongly reduced in the figure to 16 vertical pixels, marked by the upper grey area. The remainder of the CCD is now shaded and belongs to the passive area (white). Technically, the best way to achieve this is to increase the covered area during the device's manufacture and leave only the small upper area "open." Doing so makes the CCD useless for imaging applications. If variable modification is preferred, light stops, baffles, and apertures can be placed into the beam and into the spectrometer to achieve the required shadowed zone. That method is flexible but carries the danger of stray

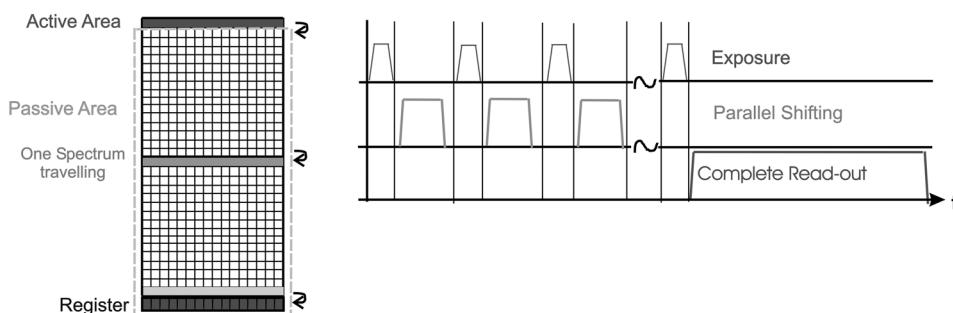


Figure 5.21 A CCD modified for kinetic measurements.

light reaching the sensor. In the example, 16 vertical pixels are shifted parallel within 25.6 μ s without loss. That time also provides the fastest repetition rate. After 64 cycles, the CCD is occupied with data, and read-out is required using the register (lower, dark-grey area). Each data set consists of 16 vertical lines binned together, shifted to the ADC, and stored. Each spectrum is stored separately. If necessary, one complete experiment goes into a 3D file. The read-out needs $1024 \times 1.6 \mu\text{s} = 1,638.4 \text{ ms}$, plus $64 \times \text{ADC time}$, which can be 1 μs .

Two problems are evident. The shortest exposure time is equal to the shift time. If the light source is not stroboscopic, vertical smear will occur, but that is fortunately not a problem in spectroscopy. The example experiment of capturing the decay of an emission source (such as a phosphor) can be conducted if the synchronization of source excitation and the start of data collection is achieved. With a pulsed and synchronized laser excitation, experiments with stimulated emission or dynamic absorption can also be performed. In any case, the number of stroboscopic experiments is limited. Therefore, kinetic systems in reality are almost always equipped with gated MCP image amplifiers. Exposure and shift are then separated, and the signal intensity is controlled by the MCP gain. Unfortunately, a different problem arises: the MCP phosphor decay time. A standard phosphor (P43) requires $\sim 3\text{--}5 \text{ ms}$ to decay to the 1% emission mark; for many experiments, that time is too long and creates time smear. A phosphor P46, operated under moderate cooling (perhaps +10 °C), will decay within 3 μs , which, in turn, fits nicely the parallel shift time. Compared to the P43, the P46 is less efficient and emits in the blue. Both effects together necessitate a hundredfold gain in the MCP. Thus, either higher noise will be recognized or an MCP with a larger channel cross-section than the common 6 μm must be used. Such plates have larger capacity and gain at the expense of spatial resolution.

With that issue solved, the dynamic range and shift time must then be balanced. If the provided dynamic range of $>65,000$ is not required, but 50,000 is sufficient, the virtual CCD programming can be used to realize, for example, a shift time as fast as 3 μs per block of 16 lines. In this section's example, all 64 single measurements can be done within 200 μs . Of course, due to the phosphor afterglow in the same time frame, the gate time must be cut into the ns regime. High-performance CCD camera systems can combine the synchronization of gate and gate delay with respect to the trigger and independent from the shift control. They also can arrange a time schedule so that the time per track can be varied during the kinetic record. Hence, within one experiment, the time, gate, and gate delay can be modified, which is beneficial for all kinds of *e*-functional behavior. It is even better if the gate and the shift can be completely controlled via external equipment.

5.8.10.2 Double-pulse measurements

The limiting factor for fast sequences with frame-transfer and full-frame CCD systems is the parallel shift time; herein lies one of the advantages of the

interline CCD. Its disadvantage of having a passive column of pixels beside each active row that divides the active surface by two and makes the optical reproduction asymmetric by microlenses also includes an immense advantage. The parallel shift (the interline transfer works in horizontal order) of all content typically happens within 0.3 μ s, as shown in Section 5.8.2 and Fig. 5.14. That fact enables two illuminations within a short interval. After the first pulse, the shift is done, and the CCD is ready for the second pulse. Once that is recorded, the read-out occurs. First, it transfers the content of the passive pixels, which is the first data set, and then it performs one more shift and reads out the second set of data. Several dynamic experiments in imaging and spectroscopy bring very useful data with only two measurements; examples include the saturation effects in flames or plasmas. Other experiments, such as fluorescence or absorbance, can reveal nonlinear results when driven by a laser. The application can compare two superimposed lasers, as used in CARS. Because strong lasers are often used to pump the sample, provisions are required to prevent the laser light from reaching the detector. Even if “only” the advantage of creating the next spectrum or image during the read-out of the other set of data is applied, the duty factor improves.

5.8.11 Extending the spectral efficiency into the deep UV

Because silicon is efficient from \sim 1100 nm to the extreme x-ray range of \sim 0.1 nm (or 0.9 eV to 10 keV), rear-side illuminated CCDs can be used below 200 nm without a covering layer. That would mean that the whole detector system needs evacuation, even when not used, to avoid contamination. Unfortunately, the high-energy rays do hurt the structures of the CCD, with increasing effects towards increasing energies. Thus, the efficiency of systems exposed to EUV and x rays drops over time. A useful alternative is to add a UV scintillator, as described in Section 5.2.5. If the lower wavelength limit is not less than 106 nm, windows made of MgF₂ or CaF₂ will save the detector from ambient energy. That technique will even work with front-illuminated detectors, such as arrays. The scintillators are available as a gel (such as barium fluoride) or foil (such as potassium silicate). The efficiency of the materials between 30–300 nm (4–40 eV) provides values between 10–90% QE. They emit between 300–450 nm. In the range of the scintillator’s emission, a slight absorption occurs, which reduces the output efficiency by \sim 10–15%. No other drawbacks have been reported.

5.8.12 NIR and IR area detectors

Like single-point detectors, NIR-IR detectors are hybrids, based on silicon arrays or CCDs. As with all NIR-IR detectors, the amount of background and noise increase towards longer wavelengths. Table 5.3, similar to Table 5.2, presents the key parameters of NIR area detectors.

The table shows the dark charge with LN cooling only, producing about -120 °C at the sensor. The parameter is defined in electrons/second (instead of e⁻/hour). The rows on the right represent the working range. The

Table 5.3 Parameters of NIR area detectors.

Device	Pixel Size [μm]	Capacity [e^-]	Dark Charge [e^-/s] at -120°C	ADC-Time [μs]	Read-Out Noise [e^-]	Linear Dynamic Range	λ_{\min} [nm]	λ_{\max} [nm]
CCD	20 \times 20 FD	20×10^4	2×10^{-5}	10	4	50,000	450	1,100
				1	12	16,000		
	20 \times 20 BD	20×10^4	0.003	10	4	50,000	200	1,130
				1	12	16,000		
UV-Vis Array	25 \times 2,500	8×10^7	0.001	10	800	100,000	200	1,100
				1	2,000	40,000		
NIR Array/Standard	25 \times 500	1×10^8	1×10^4	1	1,000	10,000	800	1,700
				0.2	5,000	2,000		
NIR Array/Extended	25 \times 250	1×10^8	3×10^6	1	2,000	1,500	800	2,150
				0.2	8,000	375		

two upper detectors are used for reference, and both are CCDs with deep-depletion technology. The first one (FD) is front illuminated, and the second one (BD) is back illuminated. Next comes the normal silicon array, and finally two InGaAs arrays with different wavelength ranges and sizes. As expected after Sections 5.6.1–5.6.8, the dark charge rate accelerates with the wavelength extension into the IR range. It squeezes the linear dynamic range quickly into values that make the application difficult for spectroscopy. The corresponding efficiency curves are part of Figs. 5.6 and 5.16. Because the bandgap of Si is near 1130 nm and no absorption happens above the bandgap, the detection of lower photon energies is realized by hybrid technology. The hybrid detector consists of a Si array or Si CCD with a superimposed detecting stack that is electrically bound to the Si elements. It can consist of InGaAs or any other 3-5 compound, which converts the photons into charge carriers and transfers them to the silicon. There, the charge is stored for read-out, as in a standard CCD. Due to the higher noise factor of the hybrid element, not only does the thermal dark charge grow but also the read-out noise, even at rather small pixel sizes.

Hybrid systems are available as line arrays or as 2D CCD systems. Both provide similar noise and dark levels. As Table 5.3 shows, fast ADC conversion is used to keep the dark charge low at sufficient light intensity. Due to the high total-noise level, there is still little dynamic reserve, as also documented in the table. Background and noise increase exponentially with the upper detection-wavelength limit. If the experiment can tolerate a background of 10% of the saturation, the 1.7- μm NIR array can theoretically integrate up to 1000 s, whereas the maximum integration time of the 2.15- μm

version would only be 3.33 s. Unfortunately, these numbers are valid for a detector looking into an environment that is also $-120\text{ }^{\circ}\text{C}$, which means that the thermal signals out of the spectrometer must be added and will drastically shorten the allowed time between two read-outs. Real-world data for the $1.7\text{-}\mu\text{m}$ NIR array will be such that the 10% limit is reached after 10–20 s, and the $2.15\text{-}\mu\text{m}$ version would be at 10% after <1 s. For that reason, no data for detectors of longer or wider wavelength ranges are considered in the table. The practical consequence is that IR area detectors for civilian applications have no real chance against modulated, single-element detection systems, such as scanning spectrometers, and interferometers with FFT data processing.

5.9 Other Area Detectors

5.9.1 CID and CMOS arrays

Charge-injected devices (CIDs) are a variation of CCD technology that features several extra functions: Each pixel can be addressed separately. The read process does not destroy the charge within a certain pixel. The charge is sensed but stays in place for further integration; to achieve this, address lines run between the pixels in both dimensions beneath the photoactive surface and pixel layer. The surface efficiency is 100% (also true for a CCD). Erasing the charge is a flash process, common to the whole device, that switches all pixels to ground. In summary, a CID allows one to sense and store the degree of saturation, for certain or all pixels, and decide whether the integration should happen. Due to the more complicated structures, background and noise are higher than in a CCD, although the principal parameters are similar.

Complementary metal–oxide semiconductor (CMOS) devices consist of a linear or area array of active surface sensors attached to discrete electronics. That arrangement allows direct, active access to each detector element. A CMOS device is also called an “active pixel sensor” (APS).

A CMOS device is a 3D structure, as illustrated in Fig. 5.22. In the top level (the active pixel area where the light arrives) are independent silicon photodiodes, marked in yellow. Similar to a front-illuminated CCD, the wiring is located between the pixels. The next level is occupied by discrete pre-amplifier/integrator combinations (blue), one each for every diode. Below that are two levels for the control and read-out lines (red and green dotted lines), allowing direct access to each amplifier/integrator. The next level is home to the driver and gate electronics. It connects to the level for the clock, timing, and input–output circuitry. The base level contains the socket, which has all required connections. CMOS devices are self-contained detectors that, once programmed, do not require complicated, external online control.

5.9.1.1 Typical CMOS parameters, and comparison to CCDs

All actual CMOS devices are front illuminated. The surface fill factor varies between 40–70%, similar to that of a front-illuminated CCD. Thus, the

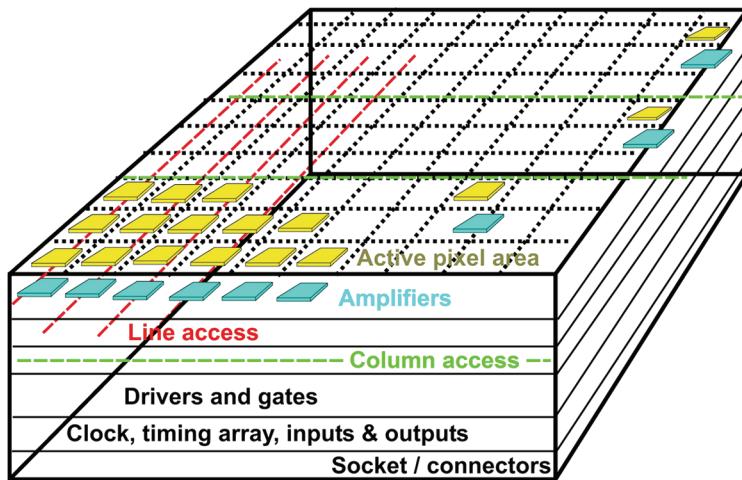


Figure 5.22 Principle of a CMOS device.

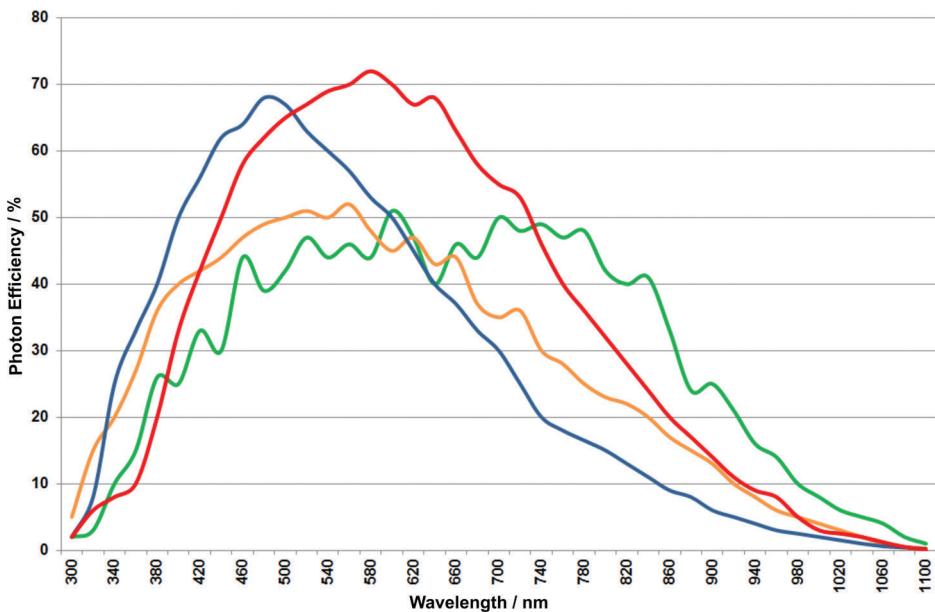


Figure 5.23 Four typical CMOS efficiencies.

efficiency curves shown in Fig. 5.23 must be corrected for the active surface, which means that a 50% QE represents the most efficient models. The front illumination also circumvents UV sensitivity. Some of the devices are equipped with microlenses (not shown in Fig. 5.22) to collect and guide more light to the diodes at the expense of spatial resolution. The pixels have a square shape, and their size reaches from $<3 \times 3 \mu\text{m}^2$ to $8 \times 8 \mu\text{m}^2$.

The majority are $\sim 5 \times 5 \mu\text{m}^2$, which offers a superb geometrical resolution that is good for photography and imaging. For spectroscopy, however, the tiny pixels are counterproductive for several reasons. The capacity is less than $50,000 \text{ e}^-$ in the very best case; for most devices it is $<5000 \text{ e}^-$. Assuming a noise level of 2 e^- , the dynamic range will be 12 bits, at best. For some emission experiments, that range will be sufficient, but for relative spectroscopy (i.e., absorption measurements) it is surely not. Looking beyond the Rayleigh diffraction limit, Eq. (2.23) furthermore shows that very small pixel sizes (below $\sim 5 \mu\text{m}$) are useless. Supposing a slit width of twice the pixel width to utilize best possible operation, the entrance slit might drop below the Rayleigh limit, and interference phenomena and ghost peaks might appear. Some of the CMOS devices allow binning of a few pixels in both axes. Unfortunately, that does not help because all pixels are separate elements, both physical and electric. Hence, binning does not create “superpixels,” as in the case of CCDs. Because each pixel has its own pre-amp, the SNR is not improved by binning data; only memory capacity is saved.

On the other hand, this technology includes a direct addressing and read-out for each pixel, and a particular read-out and storage of any x/y address, which provides the advantage of reading out selected areas or even single pixels while the others integrate the light. The read process is very fast—A/D conversions happen in the nanosecond range, enabling large numbers of full frames per second because no shift or other overhead times must be added. Because each pixel has its own “private” electronics, the background and gain can be adjusted individually, which allows for both an extreme homogenous background and full-scale alignment over the full surface (in a CCD application, this must be done externally, if required). In a scientific CMOS device, the parameters for each amplifier can be stored in the device itself, reducing the required external software support during operation.

The separated electronics provide one more advantage not possible with CCDs: the flash and gate capability. Each amplifier/integrator can be started/stopped individually, with a typical precision and resolution of $<1 \mu\text{s}$. The advantages of that capability allow extremely flexible signal treatment patterns. Unfortunately, the geometric size of an actual CMOS device is not advantageous for spectroscopy. The largest available currently have a $<15\text{-mm}$ side length, while the most popular models are $<10 \text{ mm}$. It follows that the typical field lengths of at least 25 mm, available in modular spectrographs, are only partly used, which either limits the spectral interval or bandwidth/resolution. Thus, the capabilities of CMOS devices for optical spectroscopy are very limited today; their main use is in photography and fast imaging. The technology has a promising outlook. If improved pixel and device lengths, UV capabilities, and surface efficiencies become available, CMOS devices could grow to be a serious competition to CCDs. As that is not yet the case, the remainder of this book will not discuss it further.

5.9.2 Position-sensitive detector plate

A position-sensitive detector (PSD) is a conductive plate, being loaded with a static, positive charge, and is equipped with four pulse sensors. The arrival of a pulse (a cloud) of electrons at the plate triggers an immediate drop of charge at the point struck. A ring-shaped wave will start from there and move over the whole surface of the plate. The wave will need different travel times (except if the pulse hit the center perfectly) at the four corners, which are recorded. Figure 5.24 shows two subsequent pulses, P1 (red) and P2 (blue). The right part of the graph demonstrates the arrival of the pulse wave at the different sensors and their relative amplitude. In a comparator, the four signals are correlated, with the option of double correlation: by the time of travel and by the amplitude. After the comparator has defined the position of the original pulse, it writes a count into a matrix storage. Thus, the memory contains both the frequency of pulses and the spatial distribution. It is even possible to create a time-interval (t_i) distribution curve. While the comparator is processing data and the charge recovers (together, they are called the dead time t_r), no further pulses can arrive or else it will be lost, and the dead time increases. The dead time is in the range of 10–100 ns, depending on the dimensions of the plate, which will be between several mm up to 100×100 mm. The PSD, in practice, creates no background signal, but it can be disturbed by cosmic rays. It only responds to accelerated electrons, which predisposes it mainly for high energy measurements. In combination with a MCP, it makes up a highly qualified optical counting system. A major drawback of the PSD is that it cannot integrate signals. It only counts pulses, and it permits no double pulses within the dead time. The maximum count rate for a 25-mm plate, having a dead time of 30 ns, is $\sim 5 \times 10^6$ counts/s.

5.9.3 Streak and framing camera

Streak cameras combine MCP and tube technology. The electrons created by a cathode (source) fly towards the anode (screen); they carry 3D information.

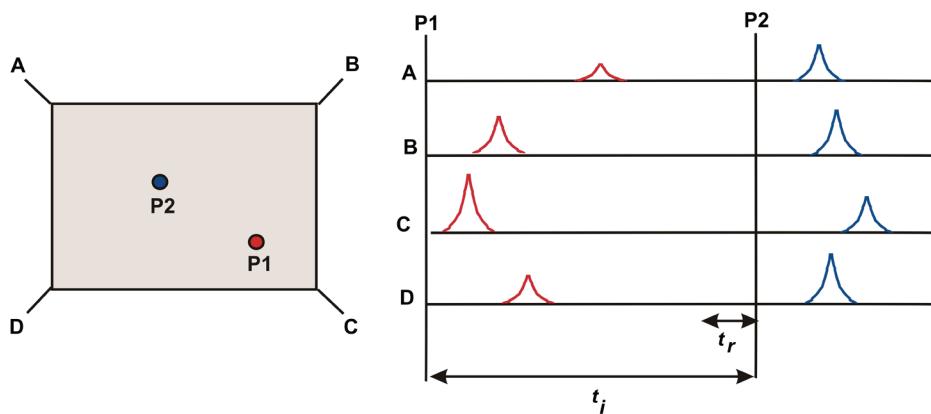


Figure 5.24 Function of a PSD detector.

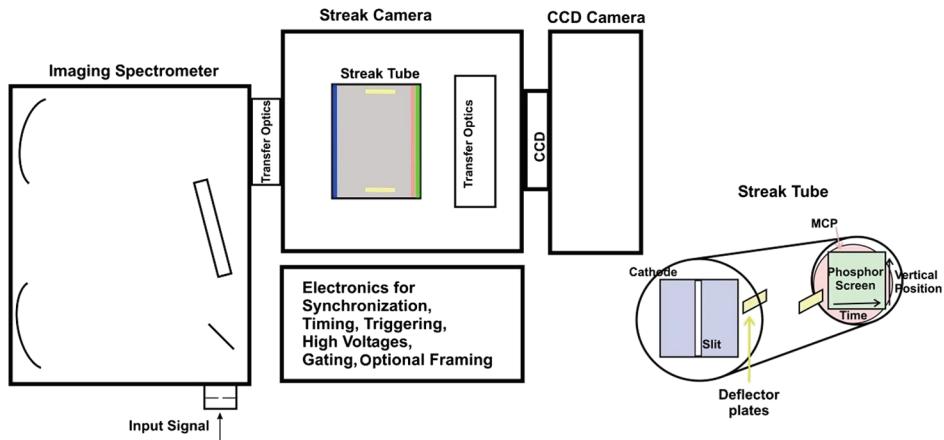


Figure 5.25 Streak spectrometer and the principle of a streak tube.

The horizontal axis represents time, and the vertical axis is the output position of the spectrometer. Intensity is the third parameter. A framing camera is, in principle, the same, but it has additional electronics to ensure that the timing is repeated after each trigger, and the system stores and overlays the signals. Streak cameras record single events that can be digitized and stored, whereas framing cameras are able to perform internal averaging after or during read-out. Streak cameras are among the fastest detection systems available today: their main applications are in the range of pico- and nanoseconds. Some models are even able to resolve femtoseconds.

The basic idea of the streak camera (Fig. 5.25) is the realization of an extremely fast, single-event oscilloscope. In the past, an electron gun delivered the electrons. After a trigger signal, the beam swept once in a pre-defined time from one side of the screen to the other, steered by the two horizontal deflector plates. The electric input-signal amplitude was used to modulate the vertical axis. An Instamatic or a vidicon camera was started parallel to the sweep and recorded the optical signal from the screen. The lifetime of the phosphor was long enough for the required illumination time. After the photo was saved or the vidicon was read-out, the system was ready for the next record.

The process became imaging and spectroscopy when the electron gun was replaced by a cathode, as in a PMT or MCP. The vertical deflector plates do not modulate the beam anymore—they are used for focusing purposes and are thus left out of the graph. In front of the cathode is a vertical slit that receives optical signals from an imaging monochromator. The two slits converge thanks to transfer optics. The vertical information can represent different positions of a sample or a light beam, or discrete arms of a fiber cable. A sweep is now dimensional, representing both the geometric height and intensity of the optical signal at the entrance slit. The result is both the time behavior of the optical signal and its distribution in the slit. If the slit is

replaced by the output field of an imaging spectrograph, 3D data can be recorded. Because the data will not be separated in the entrance, deconvolution is then required. One spectral interval (created by the grating and the slit or the plane) from several entrance positions (from fiber guides or other optical discriminators in front of the spectrograph entrance) will create a 3D signal at the screen. Right after the sweep, a CCD camera reads the screen and digitizes the content; the next sweep then follows. The intensity in the output would be very small because of the high sweeping speed, and all data need de-convolution of time versus position. To enhance the SNR, the streak tube incorporates an open MCP in front of the phosphor to intensify the signal. The MCP gain is variable and is combined to the advantage of the gating capability, which allows for more synchronization modes and the suppression of unwanted signals outside the sweep time.

After the MCP, the phosphor provides an emission signal that exists long enough to provide linear intensity for the illumination of a 2D CCD camera, thus reading out the experimental data for storage. In a framing camera, which might be the same system, extra electronics enable repeated sweeps, the presentation of which is delivered by the phosphor and triggered read-out. That process allows for the tracking data variation over time. Framing cameras also allow shifting the sweep in relation to the trigger. For repeatable experiments that last longer than the sweep time, the time tracks can be recombined in software after acquisition. The time resolution of a streak camera is $\sim 1\%$. A sweep time of 1 ns, for instance, will probably allow time slices of 10 ps to be recovered. The fastest streak cameras perform controlled sweeps in <100 ps, providing a time resolution of <1 ps.

References

1. G. F. Knoll, *Radiation Detection and Measurement*, Fourth Edition, John Wiley & Sons, New York (2010).
2. J. M. Lerner and A. Thevenon, *The Optics of Spectroscopy*, Horiba Scientific website, <http://www.horiba.com/us/en/scientific/products/optics-tutorial>.
3. Hamamatsu Photonics, *Photomultiplier Tubes: Basics and Applications*, Third Edition, http://www.hamamatsu.com/resources/pdf/etd/PMT_handbook_v3aE.pdf (2007).
4. W. Neumann, *Applications of Dispersive Optical Spectroscopy Systems*, SPIE Press, Bellingham, WA (to be published in 2014).

Chapter 6

Illumination of Spectrometers and Samples: Light Sources, Transfer Systems, and Fiber Optics

6.1 Introduction and Representation of Symbols

This chapter addresses the basics of light sources, light collection, transfer, and re-imaging. From the perspective of optical spectroscopy, light sources can be discriminated by the intensity versus wavelength distribution character. Every light source needs an optical transfer facility to illuminate a spectrometer correctly. The light sources themselves are well specified in the literature and data sheets, and thus this chapter concentrates on the transfer and special points of interest for spectroscopic applications.

The power, with respect to energy, or work of light signals is defined by the energy of the photons involved. The basics are treated in Chapter 1. Light sources emit in different ways, and thus special techniques of collection are required.

- **Laser light source:** A typical laser beam has little divergence but must not have a homogenous distribution. Different profiles over the horizontal versus the vertical plane are often found at the illuminated surface. In addition, the distribution close to the laser (near field) can differ from that at some distance (far field). Investigating both spectral and spatial variations can require beam widening and optimized illumination of the spectrometer. Laser light is generally defined by the radiation integral at a narrow wavelength interval. Lasers often come equipped with fiber optic exits, distributing their light like a fiber optic instead of the original laser.
- **Cone-shaped emission:** Cone emitters, such as LEDs, hollow cathode lamps (HCLs), or deuterium lamps, mostly show a homogenous emission in spectral and spatial terms. Some can provide a profile

regardless. The emission profile typically presents a continuous drop of intensity from the center to the outskirts of the surface, often in a cosine function. Even though some sources can provide dark spots within the field, the general emission pattern stays as described. Similar to lasers, near- and far-field effects are found with certain LEDs. Thus, within the angle of emission, the intensity and even the spectral distribution can vary. If they are matter of analysis, special setups are required. In any case, the emitted radiation power will show a continuous decrease of radiance over distance.

- **Ball-shaped emission:** The radiance coming from cone emitters and (more or less ideal) ball-like emitting sources (thermal and discharge lamps) will drop in a square function with distance. The illuminated area and its distance to the source must be taken into account.
- **Diffuse light sources:** A diffuse source is best described by the rules for point sources, utilizing the same calculation parameters. The main difference is that the source volume and the source area have a different relation.
- **Synchrotron radiation:** The synchrotron is the perfect light source: it provides collimated beams flexibility in sizing, allowing easy transfer into spectrometers. The energy distribution versus photon energy/wavelength is homogenous over the whole range of interest, from the EUV to far-IR. The amount of energy can be chosen/dimmed. Unfortunately, there are several problems: availability, running costs, portability, and the price of installation. A standard user of optical spectroscopy will rarely use a synchrotron as a light source. The collection of synchrotron radiation follows the rules presented in Section 6.4.1.

This chapter does not review photometric and colorimetric applications and parameters, only spectroscopy. In photometry and colorimetry, only the spectral range of the human eye is used, and the measured data are superimposed by weighting coefficients to adapt to the sensitivity of the human eye.¹ The specifics of spectroradiometry are treated in the complementary book on applications.²

6.2 Radiometric Parameters

Radiometric parameters are frequently encountered when dealing with light sources and optical transfer. Figure 6.1 illustrates these factors. The spectral radiant energy is the optical work of a source; its symbol is $Q_{(\delta\lambda)}$ or $Q_{e(\delta\lambda)}$, and the unit is J/nm, equal to Ws/nm. If the optical work is reduced to the optical power, what is left is the (spectral) radiant power, also called the radiant flux, with the symbol $\Phi_{(\delta\lambda)}$ or $\Phi_{e(\delta\lambda)}$ [W/nm]. Both parameters define a source without taking into account the area of origin. If the power/flux is related to the area of its origin, it is called the radiant exitance or radiant emittance:

$$M_{e(\delta\lambda)} = \Phi_{e(\delta\lambda)} / (m^2 \times nm) [W/(m^2 \times nm)]. \quad (6.1)$$

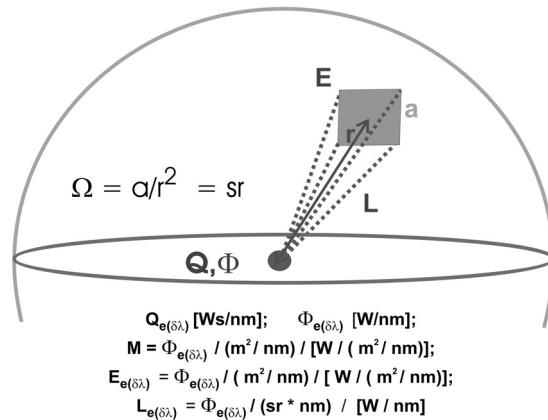


Figure 6.1 Radiometric nomenclature and parameters.

The inversion is the relation of radiant power at the illuminated area, which is a passive, illuminated area; it is called the spectral irradiance, and the equations are practically the same:

$$E_{e(\lambda)} = \Phi_{e(\delta\lambda)} / (\text{m}^2 \times \text{nm}) [\text{W}/(\text{m}^2 \times \text{nm})].$$

The SI system does not define angles in degrees or angular functions. To include the angle of emission from the source, the normalized angle Ω is entered. Keep in mind the effective aperture of spectrometers (Section 2.6.2), which is the counterpart. Ω is equally valid for reception and transfer, as well as the emission in spherical shape; indeed, it acknowledges that the density of light diminishes by the square of the distance to the source: $\Omega = a/r^2$. In the literature, the normalized angle is often called the light guiding factor.

The expression “steradian” combines the Greek word for “solid” and the Latin word for “radius”. If a geometric area a of 1 m^2 is illuminated from a spot source with a 1-m distance r , the result is 1 sr . A whole sphere consists of the wall area of $4\pi r^2$; at a 1-m radius (2-m diameter), that is 12.567 m^2 , which also represents 12.567 sr . If a light source is imagined with homogenous spatial radiation in a sphere, the radiation will be equally distributed everywhere in space, but that phenomenon only becomes apparent at the wall. If the radius increases by two, the wall area will increase by four, producing a light density of $1/4$ per unit area. In general, Ω is a very good tool for normalized calculations.

The normalized (spectral) light density is called the radiance, expressed by

$$L_{e(\delta\lambda)} = \Phi_{e(\delta\lambda)} / (\text{sr} * \text{nm}) [\text{W}/\text{nm}]. \quad (6.2)$$

The light-guiding factor into space is found by Eq. (2.15): $\Omega = a/r^2$. The density/radiance of radiation traveling fan-shaped into space is constant if Ω is constant. Parameters that contain W/m^2 are not very useful for spectroscopy—smaller dimensions are more applicable. The reduction of dimensions to $\mu\text{W}/\text{mm}^2$ produces the same numbers but can be applied better. A mw/cm^2

version, often found in the literature and also used here, uses numbers of a factor of 1/10 compared to the other two versions. The normalized radiance/spectral density is

$$L = (\Phi \times \Omega) / (A \times \delta\lambda) [\mu\text{W}/(\text{sr} \times \text{mm}^2 \times \text{nm})], \quad (6.3)$$

where L is the spectral density of the light/radiance; Φ is the light flux/radiant power of the source in W, mW, μW ; Ω is the normalized light-guiding factor of the radiator; A is the illuminated or emitting area in m^2 , cm^2 , mm^2 ; and $\delta\lambda$ is the wavelength interval involved, in nm.

6.3 Advantage of Using Ω and sr

Consider a spot light source distributing light in the shape of a cone (LED or D₂ lamp) with a 64-deg full angle and a radiant power $\Phi = 1 \text{ W/nm}$. A lens of 50-mm \varnothing can be placed at a 50-mm distance, producing $f/1$ and $\Omega = 0.785$. The source at that distance can illuminate an area 5.3 cm in diameter, or 22.06 cm^2 , which, in turn, represents $\Omega = 0.882$. The two Ω result in the ratio of $0.785/0.882 = 0.89$. The radiance $L = (1 \text{ W} \times 0.89)/(1963 \text{ mm}^2 \times 1 \text{ nm}) = 0.453 \text{ mW}/(\text{sr} \times \text{mm}^2 \times \text{nm})$. A second lens of the same diameter will illuminate the spectrometer by the required cone.

It is assumed that the aperture of the spectrometer is $f/3.57$ (i.e., $f = 250 \text{ mm}$, with a grating of 70^2 mm^2). The resulting Ω of the instrument is 0.0784. The slit area will be 3 mm^2 . Furthermore, we find a radiant flux into the spectrometer of $\Phi = [0.453 \text{ mW}/(\text{sr} \times \text{mm}^2 \times \text{nm})] \times (0.0784/0.785 \text{ sr} \times 3 \text{ mm}^2) = 0.136 \text{ mW}$. So far, so good. However, there is a shorter way if we directly ratio the Ω of source and spectrometer:

$$\begin{aligned} \Phi &= [(1 \text{ W/nm} \times 0.0784/0.882) (\text{sr} \times \text{mm}^2 \times \text{nm})]/(1963 \text{ mm}^2 \times 3 \text{ mm}^2) \\ &= 0.136 \text{ mW}. \end{aligned}$$

If only a single lens replaces the lens pair, it must have two different focal lengths and thus two different Ω ; the result would be the same. In addition, if all values and Ω are put into one single equation, the $\Omega = 0.89$ of the first lens would be removed. Note that none of the examples include any loss by reflection or absorption.

6.4 Different Types of Radiation and Their Collection

6.4.1 Laser radiation

Many lasers create almost-perfect parallel beams with cross-sections from the μm range to above 10 mm, as shown in Figs. 6.2(a) and 6.2(b). Semiconductor lasers, on the other hand, can have rather large diverging angles and cross-sections, as shown in Figs. 6.2(c) and 6.2(d). In relation to the acceptance angle of a spectrometer, (a) and (b) can be seen as collimated, whereas the divergent beams of (c) and (d) can be treated as a conical beam. To illuminate

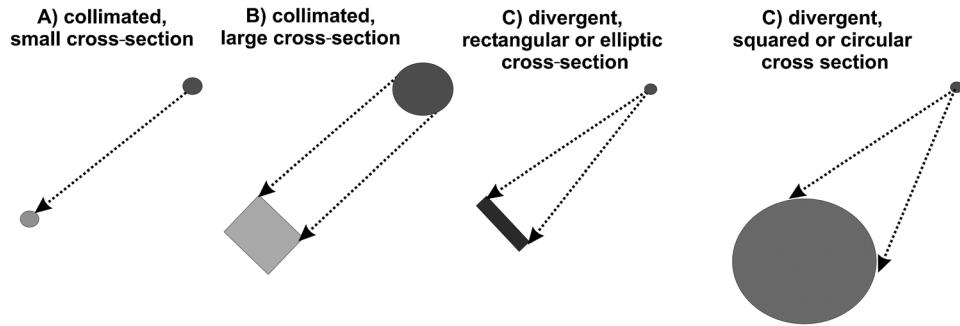


Figure 6.2 Four typical laser far-field images.

a spectrometer correctly, a cone beam of the correct angle is required in any case. Therefore, a thin, collimated laser beam requires beam widening and refocusing to illuminate a classical spectrometer. The beam cross-section can reach from $<1\text{ mm}\varnothing$ (i.e., gas lasers) up to $30\text{ mm} \times 30\text{ mm}$ (such as excimers and CO_2). The spectral bandwidth of laser lines is also extremely diversified: it can vary from $<1\text{ pm}$ (gas lasers) to $>100\text{ nm}$ (such as Ti-Sa or dye). The wavelengths reach from the extreme UV (such as electron lasers and excimers) to the far-IR (such as CO_2). The main duties of the spectrometer are the selection of a single line, the limitation of the bandwidth, and the spectral analysis. Many lasers can be obtained with an internal spectrometer designed for the required task. They use prisms and gratings, as well. The grating solutions often use wide angles between the illumination and the dispersion beam, requiring few components. Of course, one must transfer the full beam (energy), and at the same time provide the dispersion and resolution planned. If the original laser beam is divergent, its behavior must be taken into account.

Figure 6.3(a) illustrates the principle of a normal-incidence (NI) grating monochromator. A rather high dispersion is reached by a 70-deg internal angle (NI). It is of additional advantage that the efficiency concentrates on one plane of polarization and can be up to 90%. An NI monochromator needs to be optimized for the laser with which it is combined (λ , beam cross-section,

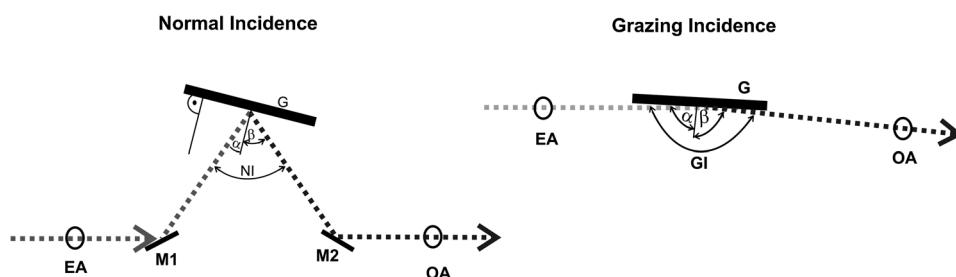


Figure 6.3 Laser-internal spectrometers: (a) normal incidence and (b) grazing incidence.

power losses at the grating, polarization). Consider the following example. The bandwidth of a 248-nm excimer is limited to 1 nm. The original beam is 10-mm high and 15-mm wide, and the polarization is horizontal (S). The entrance aperture (EA in the figure) will transfer the complete beam into the spectrometer. Diverter M1 transforms the beam to $10\text{ mm} \times 18.3\text{ mm}$, while the grating has a size of $12\text{ mm} \times 20\text{ mm}$. It is 2400 mm^{-1} and works under 14.1 deg, producing a dispersion of 0.14 nm/mm . If the output aperture (OA) is set to a width of 7.15 mm and is 500 mm from the grating, a 1-nm bandwidth will leave the spectrometer. For even higher dispersions, a grazing incidence monochromator can be used. Because the illumination angle will be re-routed less than 6 deg at the grating, the diverter mirror would probably be superfluous. Due to the very high working angle of the grating, the dispersion will be extremely high. The sum of angles α and β will be between 84–88 deg. Thus, even for a slim beam cross-section, the widening of the input beam becomes obsolete.

First, the grating would be 300 mm^{-1} and operate under 1.55 deg; the inclusion angle (GI) between the arms would be 87 deg, resulting in a dispersion of 0.25 nm/mm at a distance of 500 mm. In this example, the input beam would require a horizontal grating dimension of $\sim 280\text{ mm}$. Thus, the grating would have a width of 300 mm at a height of 12 mm. To let a 1-nm bandwidth pass, the output aperture would be 4-mm wide. A drawback of the grazing incidence setup is the rather short wavelength range covered by one specific grating. In the example, the range would be 180–320 nm, less than one octave. For all internal dispersers and optics, attention must be paid to the thermal power conversion producing the heat or ablation processes. Consequently, gratings and mirrors are often made of pure materials (steel, copper, brass, gold).

Besides the selection of spectral lines and limitation of bandwidth, external spectrometers are used for the spectroscopic analysis of lasers. The bandwidth of the measurement system should be at least 10 times narrower (30 times is better) than the required bandwidth of the lines tested. As soon as bandwidths of less than 100 pm are considered, problems arise because the measurement monochromator needs to perform on bandwidths of $<10\text{ pm}$. A spectrograph with parallel detection needs a bandwidth of 1 pm to create a measured bandwidth of 3 pm. Take, for example, the argon laser line at 514.5 nm. The measurement can be realized by, i.e., a 1-m spectrometer supplied with a 2400-mm^{-1} grating in first order. It will create a dispersion of 0.31 nm/mm and, at a $20\text{-}\mu\text{m}$ slit width, provide a bandwidth of 6.2 pm and a monochromator resolution of $\sim 3\text{ pm}$. If a camera with a $20\text{-}\mu\text{m}$ pixel width is used, the resolution will be $\sim 20\text{ pm}$. Towards the UV, both parameters will become worse; towards the IR, they will improve. For the analysis of gas lasers with bandwidths of 1 pm and less, a classical spectrometer is not a good choice anyway—interferometric systems are much more suitable.

The measurement of laser beam profiles of $>100\text{ }\mu\text{m}$ allows for good focusing on the spectrometer slit. Small-area beams need widening and

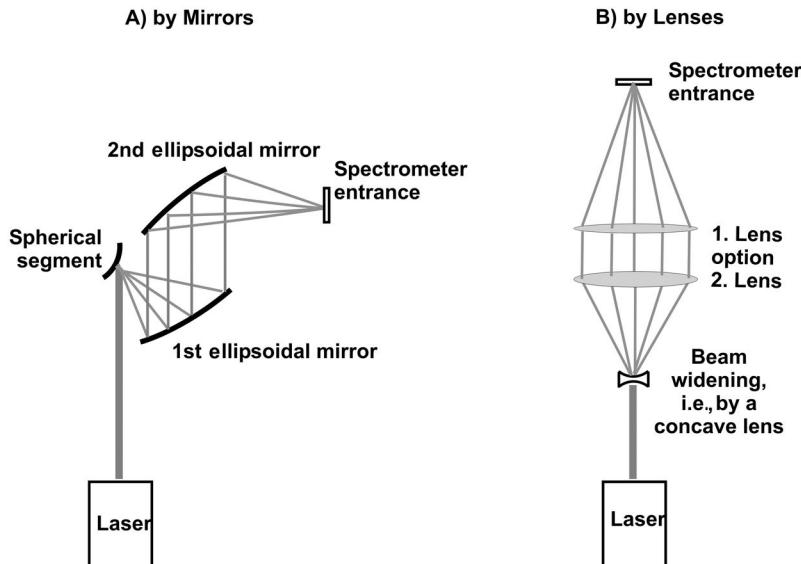


Figure 6.4 Two examples of laser beam transfer.

refocusing; several options provide that service, such as beam reflection at a ball-shaped or elliptic surface. The latter allows different reflecting angles for the two axes. The widened beam can be transferred and refocused to the slit by standard mirrors or lenses.

A convex mirror is displayed in Fig. 6.4(a), from which a cone-shaped bundle of light with known angles will travel. It will be conveyed and refocused by one or two more mirrors to illuminate a spectrometer. If the laser has just a single line, or if only lines from a small spectral interval are used, lenses can do the job, which offers linear setup solutions and symmetric beams [Fig. 6.4(b)]. If the original beam cross-section is wide enough (>1 mm), the widening can also be effected by a standard lens.

6.4.2 Cone-shaped radiation

LEDs, lamps with internal beam shaping, and some solid state lasers with rather wide bandwidths radiate into space in the shape of a cone. Thus, it is rather easy to collect the light and guide it to the spectrometer slit in the required fashion.

The left side of Fig. 6.5 portrays a linear system with lenses; the center presents an off-axis mirror system; and the right shows a typical cone-shaped radiator: the deuterium lamp. If the primary source is smaller than 1 mm \varnothing and the cone angle is narrower than 120 deg, it is possible to collect all radiation and guide it into the spectrometer. That could be realized by a single elliptical mirror, but at the expense of a nonconcentric intensity distribution in the focal spot. Based on the center figure, it is obvious that

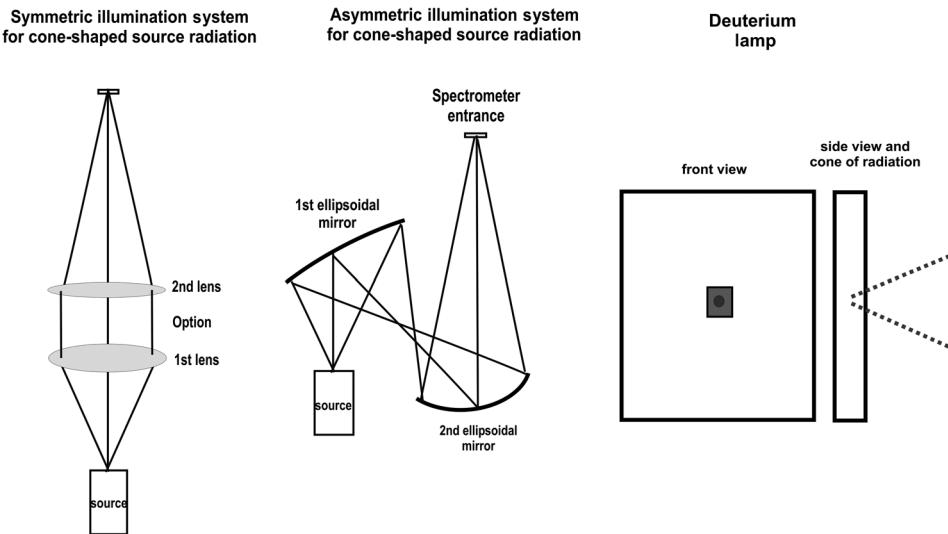


Figure 6.5 Two ways of transferring cone-shaped light.

the beams do not meet at a common point. Thus, in a two-mirror system, both mirrors must be elliptical to provide a uniform intensity distribution. If the system is made of lenses, it is much easier to reach uniformity because all rays are symmetric. A monochromator is useful to clip a defined wavelength interval and to perform spectral analysis, as well. No exceptional demands exist.

The deuterium lamp emits in conical fashion, a perfect point source with about $1\text{ mm } \varnothing$. The origin is a small discharge in a quartz envelope, the output aperture is about $2\text{ mm} \times 3\text{ mm}$ wide. Size and distance of the aperture defines shape and angle of the emission. A D_2 lamp is a wideband light source for the range of 120 to 400 nm; the general spectral behavior is part of Fig. 6.8. Hollow cathode lamps should be collected and transferred like deuterium lamps.

Consider the following light-collection example that assumes that LED, HCL, or D_2 lamp light originates from a 1-mm diameter and emits under a full angle of 60 deg. The tangent of 60 deg is 1.73. The spectrometer reviewed has an aperture of $f/5$, producing an angle of 11 deg. If the first collecting lens is set at a distance of 30 mm from the emitter, as shown on the left of Fig. 6.5, it will have a diameter of 52 mm and $f = 30\text{ mm}$. The second lens will have the same diameter at $f = 260\text{ mm}$; this setup fits both sides. Indeed, the price is an image magnification of factor of $260/30 = 8.7$. Instead of $1\text{ mm } \varnothing$, it will use $8.7\text{ mm } \varnothing$. At a spectrometer slit of up to 5 mm, not all of the light will pass. Counter calculation: If a 1:1 image reproduction is required, only a solid angle of 11 deg of the source radiation will be recovered. At first glance, that is only 18% of the light emitted, but that is only true if the density is homogeneous over the full field. In many cases, the intensity is more concentrated in the

beam center and decreases in the outskirts, often in a cosine fashion. In our example, the intensity within an angle of ± 5.5 deg stays at or above 99.5%, and drops to 87% at ± 30 deg. The central part of the radiation is relatively higher than the ratio of the areas. In addition, if the image reproduction does not use the full slit width, a nonlinear relation is given between the slit size and light flux. Parameters such as that should be considered when designing collection optics.

The radiant emittance is calculated by Eq. (6.1): $M_{e(\delta\lambda)} = \Phi_{e(\delta\lambda)} / (m^2 \times nm)$. M is defined at a certain distance from the source, with a certain full angle, or with both parameters. In any case, the required data can be recovered.

6.4.3 Ball-shaped radiation from point sources: Lamps

Thermal radiators and many discharge sources radiate like a globe or something similar. This shape makes both efficient collection and guidance into a spectrometer difficult. For all lamp radiation of a ball shape, the definitions provided in Section 6.2 are valid.

From the light ball, either only a small part of the radiation can be collected, as shown in Fig. 6.6, or the light will be lost for other reasons, as will be discussed in Section 6.4.3.4. Consider the three most popular types of lamps for the Vis–NIR range, two of which are illustrated in Fig. 6.7: on the left are two kinds of filament lamps, and on the right is an arc lamp. The deuterium lamp, also popular, is described in Section 6.4.2.

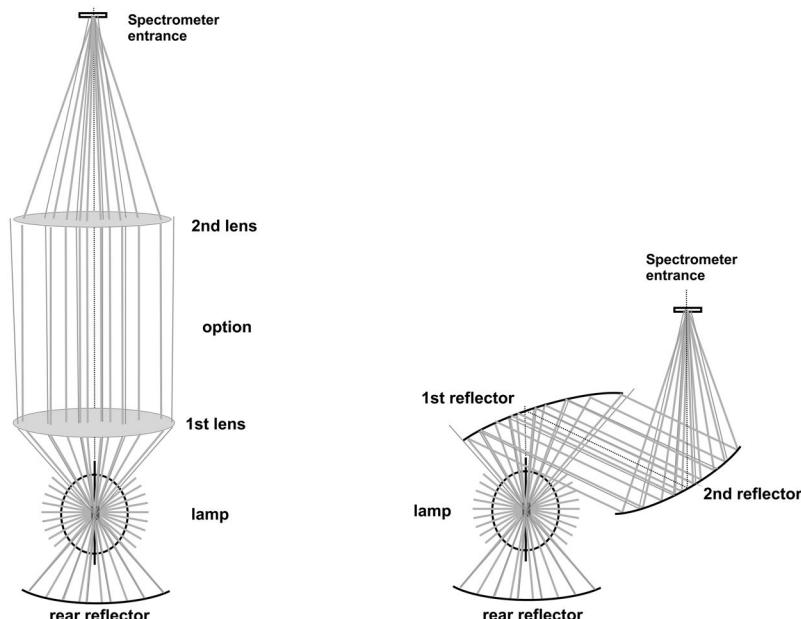


Figure 6.6 Ways to collect ball-shaped radiation and illuminate a spectrometer.

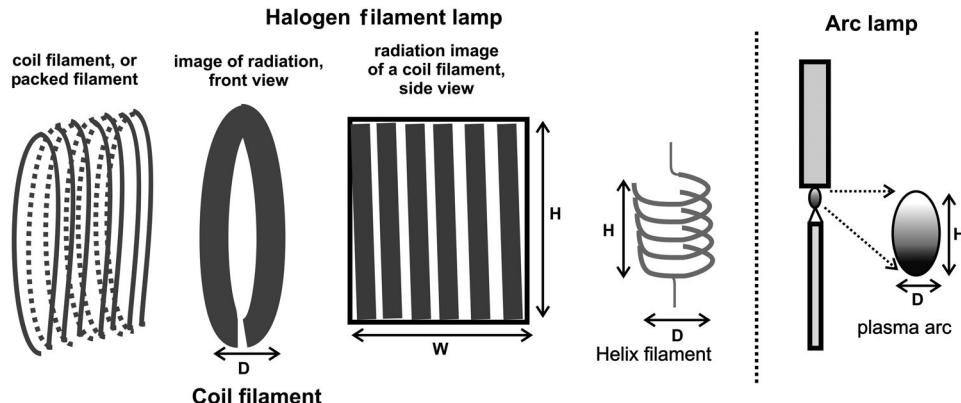


Figure 6.7 The most popular wideband radiators used in optical spectroscopy.

6.4.3.1 Thermal filament lamps

Thermal filament lamps are the most popular wideband light sources in optical spectroscopy. They are useful between <300 nm and >3000 nm. Figure 6.7 shows the typical structures of filaments. The glass envelope is neglected but will be important in the IR range. The dimensions of the filaments are as follows: D marks the width of the filament, H the height, and W the length of the package. The geometry of the filaments can appear in very different shapes. Most filament types are either packed or helix-shaped. The dimensions of packed filaments range from $W = 1.5$ mm, $H = 0.5$ mm for a 10-W halogen lamp, up to $W = 7$ mm, $H = 3.5$ mm for a 250-W halogen. Helix filaments are found between $H = 4$ mm, $D = 1.5$ mm for 50 W, up to $H = 18$ mm, $D = 5$ mm for 1000 W. The difference between a pure filament lamp and a halogen lamp is the gas inside the bulb. In a vacuum type, the heat of the filament will release molecules that will be deposited on the cooler glass wall and thus sputter it. Consequently, the transmission will reduce over time. A halogen lamp is filled with a gas that recombines the released molecules and returns them to the filament. Because the process alters the filament chemically and thermally, the spectrum is slightly modified. Therefore, a vacuum and a filament lamp do not emit identically, but they are similar enough that no further discrimination between the two is necessary in this book.

6.4.3.2 Arc discharge lamps

Arc discharge lamps are an alternative in applications that require high light density and wide spectral ranges. Xenon arc lamps reach from ~ 160 – 3000 nm, depending on the bulb material. The plasma creates an arc between the anode (the upper cylinder in the right image of Fig. 6.7) and the pen-shaped cathode. The shape of the arc is similar to an egg. The highest light density originates near the cathode and reduces towards the anode.

The size of the arc reaches from $H = 1.5$ mm, $D = 1$ mm for a 75-W standard xenon lamp [$H = 0.5$ mm, $D = 0.25$ mm for a 75-W high-density xenon lamp (at the expense of lifetime)]; $H = 2.4$ mm, $D = 0.7$ mm for a 300-W standard xenon lamp; $H = 2.7$ mm, $D = 0.9$ mm for a 450-W standard xenon lamp; and $H = 3$ mm, $D = 1$ mm for models with 1-kW and 1.6-kW xenon power. Similar dimensions apply for combined types that comprise xenon and a second gas (such as argon) or metal (such as mercury).

6.4.3.3 Spectra of the various lamp types

The spectrum depends on several parameters, but the curves presented in Fig. 6.8 are average representations. The power curves of typical lamps: 30-W D₂ (purple curve), 100-W Halogen (red curve), and 75-W Xe. The Xe is drawn in two versions: the green curve is “ozone free,” and the blue line is the deep-UV. The radiance is presented in nW/(sr × cm² × nm). Note that the spectral structure of xenon sources is far more diversified than displayed. More details are shown in the measured curve of Fig. 6.19.

Published emission curves of D₂ lamps are often truncated below 200 nm or shown fading towards the VUV. If that represents the true behavior of a lamp, it does not depend on the molecule but on the window material or the pureness of the filling. The real output curve of highly pure deuterium increases almost linearly towards 120 nm, requiring MgF₂ or BaF₂ windows. Some overshoots are evident in the range of 150 nm and at the primary line of 122 nm. Tungsten halogen sources are mainly thermal emitters that follow Planck's law. Consequently, the curve changes with the filament temperature. The curves in data sheets often represent ~3000 K. Near 325 nm, the radiant flux of comparably powered D₂ and halogen lamps are identical. Below 280 nm, even strong and hot

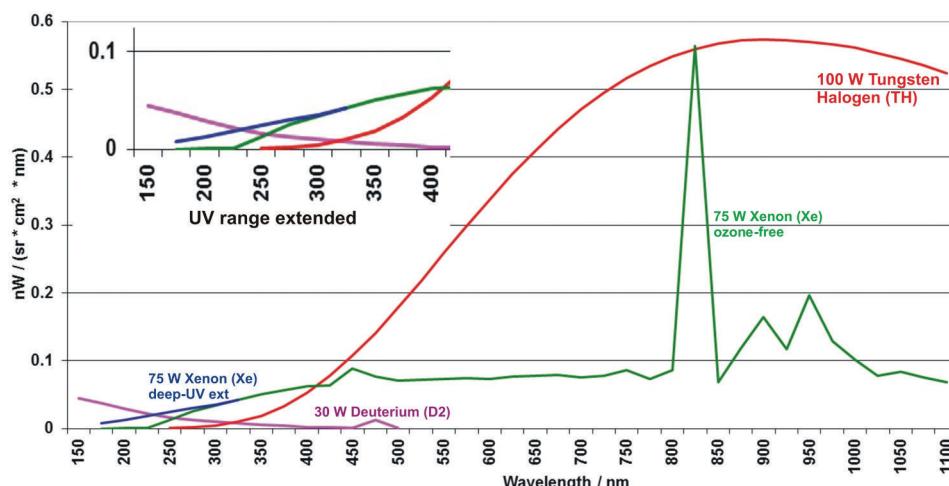


Figure 6.8 Typical, simplified spectra of deuterium (D₂), tungsten halogen (HLX), and xenon high pressure.

thermal sources do not emit enough light for practical use. Deuterium lamps are offered up to 200 W, and the useful wavelength range reaches up to ~ 380 nm. At 473.4 nm, D₂ has a strong single line, which can cause a disturbance but is also useful for λ calibration. Arc lamps are available in many different versions. Xenon has the advantage of being rather continuous over wide ranges. The useful range begins at 155 nm (special UV quartz), 180 nm (standard quartz), or 240 nm (ozone free); the last two are shown in Fig. 6.8. The plasma lines can be unpleasant—they appear between 450–500 nm and between 800–1200 nm.

In excitation spectroscopy, if the excitation wavelength is kept constant, the power overshoot at the lines of up to a factor of 100 can be of great advantage compared to the surrounding continuum. If high UV intensity is needed, Hg-Xe plasma lamps can be used, but they offer no useful continuum because of the numerous mercury lines. In principle, all plasma emitters with mixed materials are line emitters. The lines grow together due to the high temperature and pressure applied. The interpretation of spectra with wavelength variation is extremely difficult for Hg emitters. For that reason, and because it would require finely resolved curves, the display of Hg-Xe is not shown. Light sources >1000 nm are discussed in Sections 6.4.5 and 6.4.6.

The radiant flux of a light source in spectroscopic applications must not be the substantial factor. The need to guide the light efficiently into the small entrance of a spectrometer makes the light density (radiance) at least of equal importance. It substantially depends on the ratio between the power and size of the primary source (filament, arc, discharge) and is defined by the radiant emittance M (see Section 6.2, Fig. 6.1). Table 6.1 compares some often-used source types.

6.4.3.4 Light collection and transfer into a spectrometer

The intensities in this chapter are mainly presented as values of radiance [mW/(sr \times cm² \times nm)]. Figure 6.6 shows that only a section of the light produced

Table 6.1 Parameters of several wideband light sources.

Lamp Type	Radiant Medium	P [W]	Source Area $x \times y$ [mm]	Emittance M [W/(mm ² \times nm)]	Wavelength [nm]
D-30	D ₂	30	1	3	200
D-200	D ₂	200	1	20	200
XBO-75	Xe	75	0.25 \times 0.5	100	555
HBO-100	Hg-Xe	100	0.25 \times 0.25	425	555
200 W Xe-Hg	Hg-Xe	200	1 \times 3	150	555
QTZ-300	Xe	300	0.75 \times 1.5	100	555
XBO-450-OFR	Xe	450	0.9 \times 2.7	88	555
QTZ-500	Xe	500	0.3 \times 0.3	850	555
XBO-1000-W	Xe	1000	1.1 \times 2.8	150	555
JQ-50	WH	50	1.5 \times 3.2	2	555
JQ-250	WH	250	3.5 \times 7	2	555
FEL-1000	WH	1000	7 \times 20	2	555
Cesiwid-150	Globar	150	4 \times 25	0.5	1000
SiC-50	Globar	50	3.8 \times 5	0.1	1000

by a lamp will be collected if lenses or mirrors are used. Industrial lamp systems use collection optics with an aperture near $f/1$. If the first lens has a focal length of $f = 50$ mm, it will have 50-mm \varnothing , producing an active area of (rounded) 20 cm². Because the light ball of a lamp at a 5-cm distance illuminates an area of 80 cm², only 3.14 sr (25%) are gathered. The rear mirror, as in the figure, gathers another 25%. Unfortunately, that light is not used effectively because the reflected rays, which travel through the center of the source, will be absorbed in the filament or plasma. It is also counterproductive that the reflected rays travel through the lamp bulb and become refracted and diverted; they will not hit the next optical element (the first lens) under optimal angles. In reality, no more than 40% (~ 5 sr) of the available light will be used.

The arctangent of $f/1$ is 45 deg. In a spectrometer with an $f/5$ aperture, the acceptance angle is only 11 deg. To fit both angles, as shown in the left part of Fig. 6.6, the second lens also has 50-mm \varnothing but at $f = 250$ mm. The result will be an image enlargement with a factor of 5. If the reproduction of the primary source has a size smaller than or equal to the spectrometer slit, all of the light collected will be transferred at the expense of the bandwidth and resolution. In most cases, even more light will be lost, either on the collection side or the refocusing side. Light sources of interest emit rather homogeneously in all directions; only internal mechanics and the connectors reduce the perfect ball shape. Thus, the power collected will be close to the geometrical share viewed by the optics. The share of power transferred into the spectrometer will nearly linearly increase with the slit width.

6.4.4 Diffuse radiation collected by integrating spheres

Diffuse radiators include fluorescent lamps (often tube-shaped), scattering surfaces, and similar sources. Collection and transfer of diffuse light will be realized either by (a) waiving a substantial share through a limited collection area—there, the collection calculation can be treated like that of LEDs or lamps—or (b) collecting as much light as possible, taking into account that “channeling” it into the spectrometer will cost light. In the latter case, the integrating sphere is the right choice. The sphere’s inside is coated with a highly reflective yet scattering material and has a minimum of nonreflecting areas (ports). Some baffles will avoid the direct transfer of light from the entrance to the exit.

The left image in Fig. 6.9 illustrates a sphere used to collect diffuse light from large areas. Its goal is to deliver the optimal amount of radiation from the source into the sphere (grey rays). A fiber optic cable collects the light at the output and defines the viewing solid angle of collection (dotted cone, δ_L). The baffle between the entrance and exit prohibits direct transfer; the rays must reflect at least twice before they leave the sphere. The fiber cable can include a transformation of the cross-section, shown here as circular on one side and rectangular on the other, as marked by the two patterns.

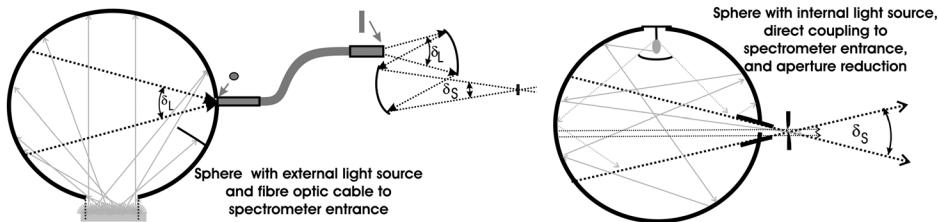


Figure 6.9 Two typical applications of integrating spheres in optical spectroscopy.

The spectrometer side of the fiber cable will fit the spectrometer best if it either has the same Ω as the spectrometer or uses conversion optics in between to fit δ_L and δ_S . If the light travel is inverted, the same setup is used for homogenous illumination of a rather large area.

The right part of the figure shows a light source placed inside the sphere. If the sphere has a diameter of <0.5 m, the source will, in most cases, be inserted from the top. Direct transfer is blocked by the curved baffle underneath the source. In larger spheres, the source is most often placed at the center of the sphere to provide even better uniformity. Metallic constructions endure strong rays and heat emitted by the source. The sample, baffles, and source holders have the same coating as the sphere wall. If the sphere is connected directly to the spectrometer entrance, it should be noted that the exiting light is diffuse, but it might take a conical shape if either the sphere exit includes some funnel alike baffles (as shown) or if the distance only allows the rays under correct maximum angle to enter the spectrometer. Regardless, the spectrometer illumination angle (Ω) and shape will fit. Unfortunately, even with the best fit, the cone of light traveling into the spectrometer will include a small percentage of parallel rays (marked by the thin dotted parallel lines), which will reduce the spectrometric performance. Fortunately, their share will be rather small, and if a double spectrometer is used, they will not reach the second stage.

Figure 6.10 presents efficiencies of often-used coatings. In the UV and the visible range, special PTFE (polytetrafluorethylen) versions are popular. Although the name “Spectralon” is a protected brand name of one of the companies in the sphere branch, it is nevertheless used as a general term for PTFE products. The PTFE-S version offers the highest efficiency in the range between 200–2100 nm, but it is more hygroscopic and angle-dependent than the PTFE-L version. Barium sulfate (BaSO_4), which is also hygroscopic, reaches deeper into the UV. In the NIR and IR, gold is unbeatable, and it is absolutely inert. All curves are shown for illumination angles up to 15 deg. Increasing the angles will lower the reflectance efficiency.

The efficiency of a sphere (besides the material constants and the illumination angle) strongly depends on the sphere diameter and the sum of the areas lost by ports. The smaller the percentage of port area is, and the

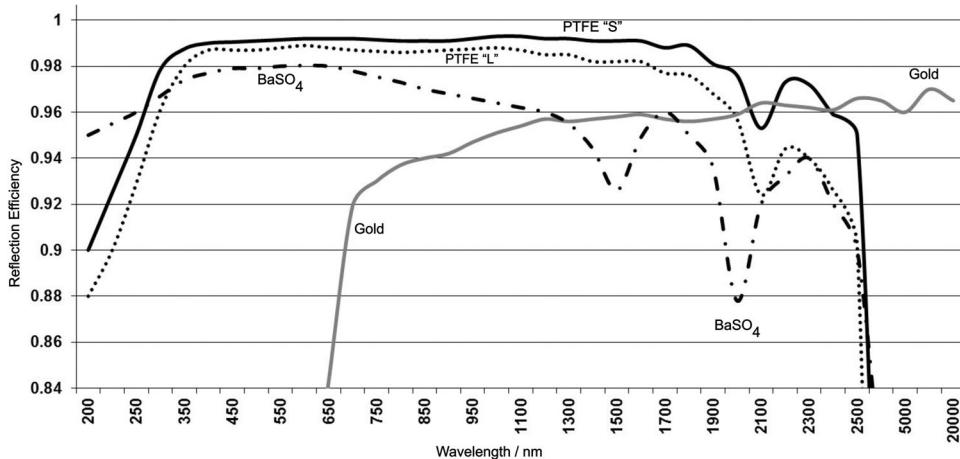


Figure 6.10 Efficiency curves of popular coatings for integrating spheres.

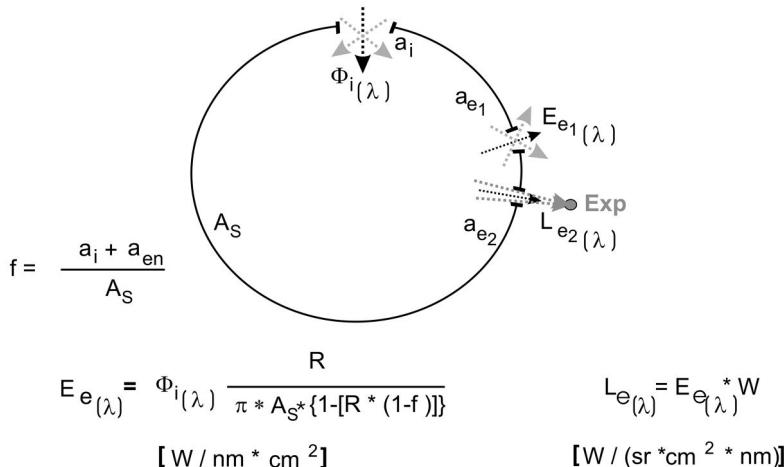


Figure 6.11 Sphere equations with graphical support.

larger the diameter is, the better the homogeneity of the radiation inside the sphere and the ports. It is a common goal to keep the total area of nonreflecting areas in a sphere <3%. The algorithms required to calculate the functions of a sphere are presented in Fig. 6.11.

The main parameters are the radiant flux $\Phi_{e(\lambda)}$, the irradiance $E_{e(\lambda)}$, and the radiance $L_{e(\lambda)}$:

$$E_{e(\lambda)} = \Phi_{e(\lambda)} \times \{R / [\pi * A_s * \{1 - [R * (1 - f)]\}]\} \text{ in } W / (cm^2 * nm); \quad (6.4)$$

$$L_{e(\lambda)} = E_{e(\lambda)} \times \Omega \text{ W / (sr * cm}^2 \times nm\text{)}. \quad (6.5)$$

Both parameters result from the radiant flux $\Phi_{i(\lambda)}$ available inside the sphere. It makes no difference whether the radiation originates inside the

sphere, or is transported into it. The following parameters are found: a_i is the entrance of the sphere and its area, a_{en} is the exit(s) and the sum of area(s), A_s is the inner surface area of the sphere, f is the ratio of all port areas divided by the inner surface area of the sphere, R is the coefficient of reflectance at the wavelength of interest, Φ is the radiant flux inside the sphere or brought into the sphere, E is the irradiance in the exit port of interest, and L describes the radiance outside the exit port, defined by the angle of divergence [it defines the radiance after the light has left the sphere, i.e., the acceptance angle of the receiver (Exp) at the exit]. Ω is the normalized light-guiding factor. All parameters apply for spectral and integral definitions.

Consider the following example: A sphere with a 15-cm inner diameter (707 cm^2) has two ports (entrance and exit), each with a 5-cm diameter ($2 \times 20 \text{ cm}^2$, or 5.544% of the inner surface); it is used to illuminate a spectrometer. The efficiency is 0.975, and the optical power entering is 1 W.

First, calculate the parameter of ratio f : $f = 20 \text{ cm}^2 / 707 \text{ cm}^2 = 0.0554$. Next, apply Eq. (6.4) to find E_e :

$$\begin{aligned} E_e &= 1 \text{ W} \times 0.975 / (3.14 \times 707 \times \{1 - 0.975 \times [0.975 \times (1 - 0.0554)]\}) \\ &= 2.54 \text{ mW/cm}^2. \end{aligned}$$

If that irradiance is guided to a spectrometer with $\Omega = 0.04$ ($f/5$), Eq. (6.5) is applied to find L_e :

$$L_e = E_e \times \Omega = 2.54 \text{ mW/cm}^2 \times 0.04 = 101.6 \mu\text{W/cm}^2.$$

Assuming that the spectrometer has an entrance slit 10-mm high and 1-mm wide, a power of $10.16 \mu\text{W}$ will travel into the spectrometer.

6.4.4.1 Collecting lamp radiation

Small sources can be placed inside the sphere, as shown in the right image of Fig. 6.9. In such a case, the sphere and its coating must not be damaged by the amount of light available at the outputs that was exercised. (Losses during coupling and transfer are described elsewhere in this chapter.) It is feasible that more light enters the spectrometer than proposed by conservative calculations.

A practical solution is a rather small sphere (Fig. 6.12) with an external 100-W halogen lamp (black rectangle), and an external 30-W deuterium lamp (dark-grey circle, meant to be placed from a side entrance under 90 deg). The sphere is directly mounted to the entrance of the spectrometer. The illumination angle is defined by the aperture in the wall of the sphere and its distance to the slit. An optional fiber optic device can either deliver calibration light or monitor the integral of the light inside the sphere.

6.4.4.2 Approaching the parameters of a sphere

The average number of reflections increases with the inner diameter, and therefore the efficiency drops. The ports can be closed by caps that have the

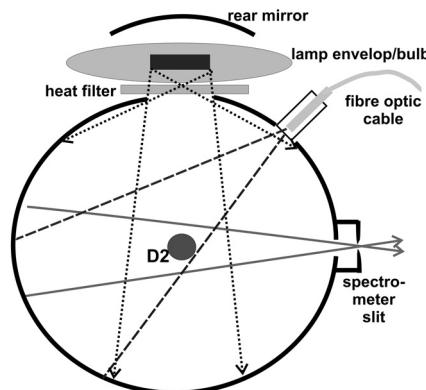


Figure 6.12 An efficient solution for light collection from large and hot lamps.

same material as the sphere itself; this allows planning with more ports than basically required, thus facilitating future applications that require one or two additional ports.

Consider the following approach for optical spectroscopy: After the number of ports is defined, the port sizes must be determined. Each port can be different, and no port should be larger than really needed. If the optimal light from a source, for example, can be transferred by a 20-mm diameter port, that value is fixed. The connection to the spectrometer also does not need to be larger than the slit height; 10 mm might be sufficient. The next step involves the slit–port distance because it defines the effective angle that will fit the spectrometer’s Ω . If optical fibers or fiber cables are applied, the ports can become rather small, and eventually a 5-mm diameter will fit. In Fig. 6.12, a 10-mm port is added for a deuterium lamp. All of the ports together result in an area of $(3.14 + 2 \times 0.79 + 0.2) \text{ cm}^2 = 4.92 \text{ cm}^2$. If that area must be kept below 2% of the sphere’s surface, the sphere should have a 245-cm^2 inner surface, which is already achieved with an 8.85-cm diameter. For this example, 10 cm is selected, resulting in an inner surface of 314 cm^2 . Next comes the coating in correlation with the light sources and wavelength range. If the desired range is 200–1100 nm, and a deuterium and a halogen lamp are available, a BaSO coating might be best. D₂ lamps typically provide a cone of 60 deg, so the proposed 5-mm port will fit. The halogen lamp will be placed in front of the 20-mm port (the spectrometer port was discussed earlier). It is a bonus that a sphere such as this can be used for several different applications.

Example of calculating efficiency

Calculating the radiant flux entering the sphere can be achieved by the lamp supplier data and Eqs. (6.1) and (6.2). The flux and irradiance available at the exit ports can be found with Eqs. (6.4) and (6.5). The

Table 6.2 Transfer and output power of a sphere (example).

Source	Wavelength [nm]	Φ at the Entrance [$\mu\text{W}/\text{nm}$]	Efficiency Factor	Φ at the Exit [$\mu\text{W}/\text{nm}$]	Total Φ at the Exit [$\mu\text{W}/\text{nm}$]
30-W Deuterium	200	4000	0.950	200	200
	250	2000	0.960	110	110
	300	1500	0.971	80	83
	350	1000	0.975	60	73
	400	500	0.978	30	80
100-W Tungsten Halogen	300	50	0.971	3	83
	350	200	0.975	13	73
	400	750	0.978	50	80
	500	2000	0.980	140	140
	600	4000	0.982	290	290
	700	5000	0.974	340	340
	800	6000	0.972	380	380
	900	7000	0.967	430	430
	1000	6500	0.965	410	410
	1100	6000	0.963	330	330

calculated sphere has a 10-cm diameter and a BaSO coating; four ports sum to a 3.15-cm² area. The result of irradiance in the ports is normalized to 1 cm² and a 1-nm bandwidth each.

The deuterium lamp, emitting a median of 2 mW/nm under 60 deg, will be mounted to the wall of the sphere. From the ball emission of the 100-W halogen lamp, a 60-deg cone can be guided into the sphere, resulting in a median value of 5 mW/nm. Expected values are shown in Table 6.2.

If both sources are turned on, the flux of both will add inside the sphere, which, in the overlapping wavelength range, sums the two values (the total Φ in the exit). One more step is required to calculate the radiant flux Φ_e at an experiment, fiber, or entrance of a spectrometer: The values of Φ at the output of the sphere are multiplied with the Ω of the external receiver to produce the radiance L_e , which is finally multiplied by the active external area to find the valid external Φ_e .

6.4.5 NIR radiation

The main sources for the range of 800–2600 nm are filament and halogen lamps. Their spectral flux differs primarily because pure thermal incandescent lamps run at higher temperatures, whereas halogen types have smaller filaments and longer lifetimes. Because both models provide their maximum flux near 1 μm (shown in Fig. 6.13), they offer the highest intensity in the NIR.

The hotter the primary source (filament) is, the higher the yield. With increasing temperature, the maximum peregrinates towards shorter wavelengths. Thus, varying the temperature allows the alignment of the amplitude behavior over wavelength to fit the detector efficiency and other parameters at the expense of the average magnitude. Because the efficiency of semiconductor detectors generally increases towards longer wavelengths, in order to quickly

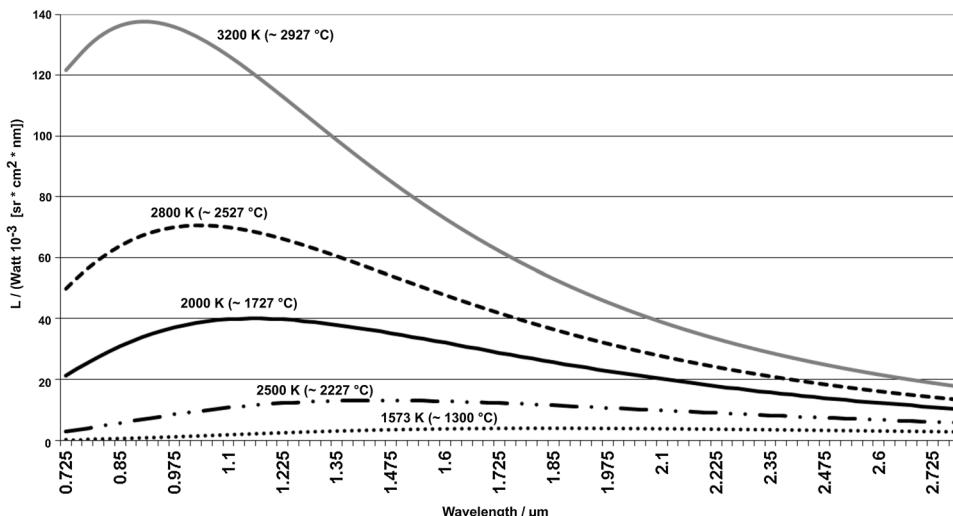


Figure 6.13 Radiance of thermal sources in the NIR at filament temperatures between 1300–2927 °C.

drop while above maximum, the lamp temperature can be used to homogenize the total system response curve. The calculations of flux and radiance are described in Section 6.2 earlier.

6.4.6 IR radiators

Tungsten lamps, at first glance, appear to be convenient for the IR, but there is a problem: The primary source, the filament, is placed in vacuum or a protective gas. The whole system is paled by a glass or quartz bulb. The optimal transmission range of quartz lies between 150–2800 nm, and that of glass at best covers 280–3500 nm. Above the upper limit, no radiation comes from the filament; the bulb itself becomes the radiator. It is far cooler than the filament, and thus it is called a secondary source. Fortunately, there are ceramic materials, such as zirconium oxide (ZrO_2) and yttrium oxide (Y_2O_3), that can operate in air and at high temperatures, providing useful IR sources. IR sources based on ZrO_2 and Y_2O_3 were developed in the 1890s by Walter Nernst. These “Nernst radiators” reach temperatures up to ~2000 K. Alternative sources in the IR are made of silicon carbide (SiC), with a maximum temperature of ~1300 K. The elements are available in very different shapes and levels of power consumption, from 0.1 mm × 7 mm to 7 mm × 20 mm. They are available in the shape of a bar and known as Globars (a derivation of “glow bars”).

The radiation follows an almost-perfect ball shape. The primary collection can be accomplished as shown in Fig. 6.6. The materials used for transmission

and reflection are determined by the required spectral range. Among them are MgF_2 (0.12–10 μm , high transmission, soft), ZnSe (0.6–15 μm , useful transmission, temperature resistant), sapphire (0.15–5 μm , high transmission, hard, temperature resistant), Ge (2–20 μm , low transmission, temperature resistant, relatively cheap), and others. The most-used reflecting materials are aluminum (<0.2 –20 μm) and gold (0.6 to $>30 \mu\text{m}$), which provide the widest working ranges.

Figures 6.14 and 6.15 show the radiance L of IR sources at some typical temperatures; they present the same curves, but the former is portrayed

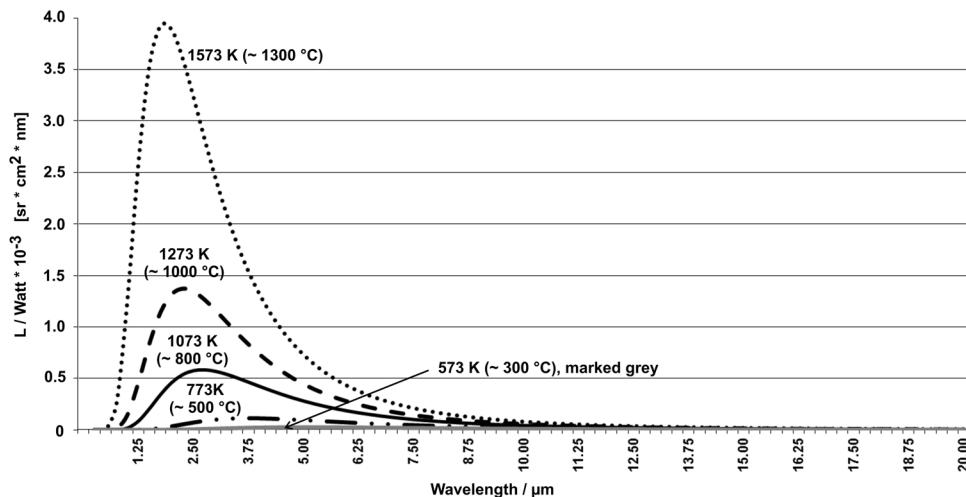


Figure 6.14 Typical performance data of thermal IR radiators, shown in a linear fashion.

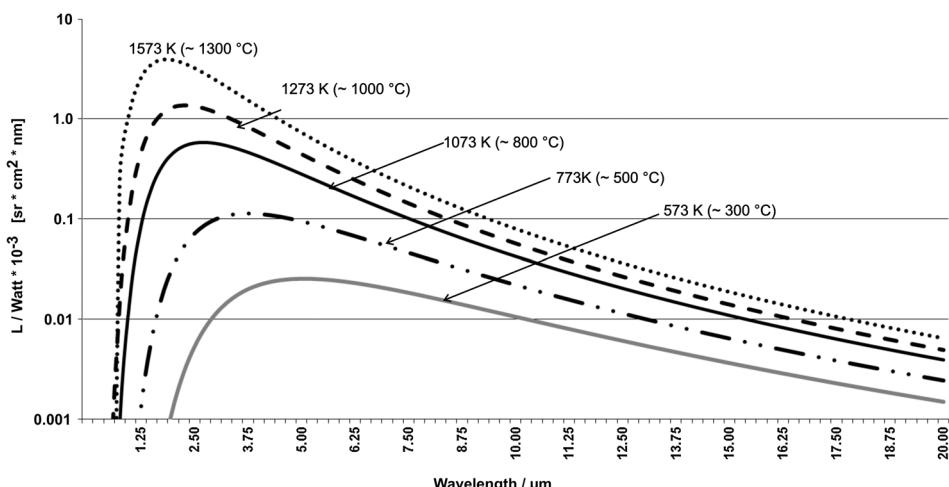


Figure 6.15 Typical performance data of thermal IR radiators, displayed along a logarithmic axis.

linearly and the latter logarithmically. It might be helpful to align the source temperature of a spectroscopic IR system with the other components in the chain. With regard to the efficiency curves of IR detectors (Section 5.6, Fig. 5.6), it becomes evident that the emission curve of thermal radiators and the IR detector curves are almost completely opposite each other, providing a rather homogeneous summing curve over wide spectral ranges.

6.5 Examples of Optimizing Spectrometer Systems

An illumination system that contains a spectrometer synthesizes its output spectral character from a number of discrete functions: the light source itself, the collection optics, the transfer system, the components of the spectrometer (such as slit size, mirrors, grating), the experiment, and the detector. This section looks at the spectrometer, with the grating as its most critical component. If, for example, a measurement system for transmission/reflection spectroscopy between 200–900 nm is under investigation, the lamp curves of Fig. 6.8 alone reveal the inherent difficulties. If D₂ and halogen lamps are combined, there will be an intensity problem between 300–350 nm. If a xenon lamp is used, there will be interference from the strong superimposed lines, making the normalization of data and spectra a difficult job, especially above 750 nm. The general drop of intensity <250 nm can probably be compensated by clever grating selection.

6.5.1 Optimization of gratings

Figure 6.16 demonstrates the optimization of gratings. A 0.5-m monochromator with three gratings on a rotating table, a D₂ lamp (solid curve), and an S-20 PMT detector was optimized for the UV range. Three measurements with different gratings were recorded, and all other parameters were kept constant. The dotted curve represents a holographic grating of 600 mm⁻¹, blazed to 3.5 deg (200 nm in first order). The dashed curve shows a ruled grating of 1200 mm⁻¹, blazed to 8.5 deg (250 nm in first order). Because the slit width was the same, it provides almost the same maximum efficiency as the “dashed” version but with a flatter top. The solid curve was taken with a grating of 600 mm⁻¹, blazed to 8.5 deg (500 nm in first order), which behaves in third order like a grating of 1800 mm⁻¹, blazed to 166 nm. Even at the beginning of the air absorption, it provides very useful intensity to 180 nm. Of course, the third-order mode of operation required extra filtering. The peaks between 185–200 nm are oxygen-absorption bands; the system was neither purged nor evacuated. The rather high lamp power created ozone in the beam, which is responsible for the peaks around 230 nm. The mirror coatings were standard for wide-range applications (Al + MgF₂). To reach a good efficiency at acceptable signal amplitude curves, an alternative might use gratings at higher orders, as the black curve demonstrates. Indeed a short-pass filter with a 250-nm edge is required to avoid order overlap.

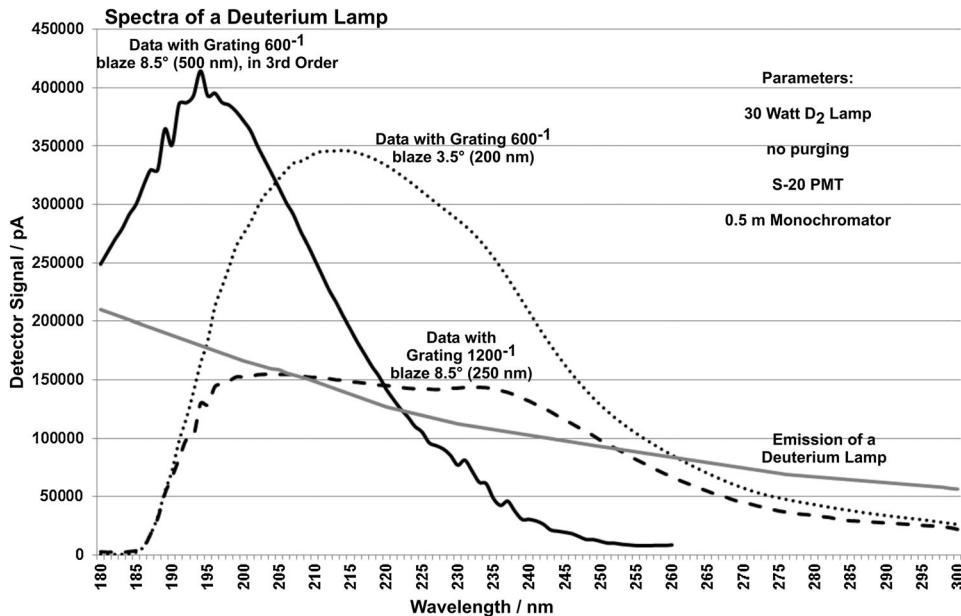


Figure 6.16 Experimental setup with a deuterium light source.

6.5.2 Change-over wavelengths of lamps, gratings, and detectors

In Fig. 6.17, a 30-W D₂ lamp and a 150-W Halogen lamp are combined in one housing. The D₂ lamp is used with a ruled grating of 1200 mm⁻¹, blaze 250 nm (8.5 deg)—shown in the left curve. The right curve represents halogen lamp running with a second ruled grating of 1200 mm⁻¹, blaze 500 nm (17 deg); the detector is a PMT (S25). This setup allows for good data up to 900 nm with only two gratings. The optimal turnover for both the lamps and the gratings is 335 nm. The signal (not the optical power) stays within a ratio of 1:20, providing an SNR variation within a factor of 5 over the whole spectral range.

More options of optimization include the following:

- The optical power of the light sources can be varied, e.g., by changing the temperature of the tungsten lamp;
- With a PMT detector, the high voltage (HV) can be reproducibly varied at pre-programmed wavelengths to adjust the signal in wide ranges. This adjustment remarkably changes the SNR and can be used for balancing. Solid state detectors do not offer this opportunity.
- If the application permits it, varying the bandwidth (or resolution) can be an efficient tool because doubling the slit width doubles the bandwidth but increases the signal fourfold (which is again a factor of 2 for the SNR).
- To keep the SNR close to the same value over the full working range, an elegant option involves varying the integration time per data point; this leads to a fourfold increase of time if the SNR needs to improve by a factor of 2.

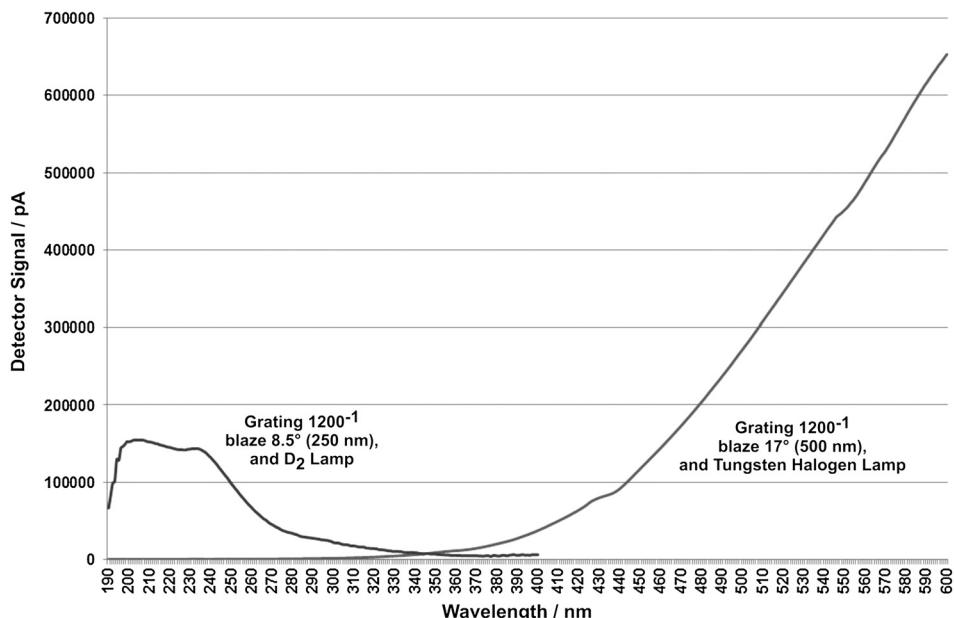


Figure 6.17 Defining the change-over point of a lamp and grating.

In the example shown in Fig. 6.18, only one light source, a 50-W halogen lamp, and a single grating (300^{-1} , 17-deg blaze, or $2 \mu\text{m}$) were used. For the change between Si and InGaAs-1700, only a short interval of ~ 50 nm is really good, whereas the change between InGaAs-1700 and InGaAs-2600 can be performed in a range of ~ 100 nm. The sharp negative peaks near 1400 nm and between 1800–2000 nm are due to water in the air; the system was not purged. Due to the efficiency behavior of the single grating, all detectors provide rather low signals below wavelengths of $1.2 \mu\text{m}$.

In Fig. 6.19, the source was an “ozone-free” 75-W xenon lamp that emits almost no signal below 230 nm. Light collection was realized by quartz lens optics with $f/1$ (lamp side) and $f/4$ (monochromator side), a 0.3-m monochromator with three gratings, and a pair of UV-Si and InGaAs-1700 detectors at the two exits. Detector changeover was programmed to 950 nm. All three gratings were 600 mm^{-1} ; the first-order blaze wavelengths are mentioned in the figure. The first grating change was at 520 nm, and the second was at 1030 nm—both points provide “flat” intensities for easy control of the fit. The bandwidth was 1 nm, ± 0.3 nm, throughout. The required order-sorting filter changes were also pre-programmed. The strong xenon lines towering over the continuum are well recorded; they make all kinds of normalization, and the creation of ratios, very difficult. In the graph, the highest peaks between 850–1000 nm have been capped on top to allow a good justification of the continuum.

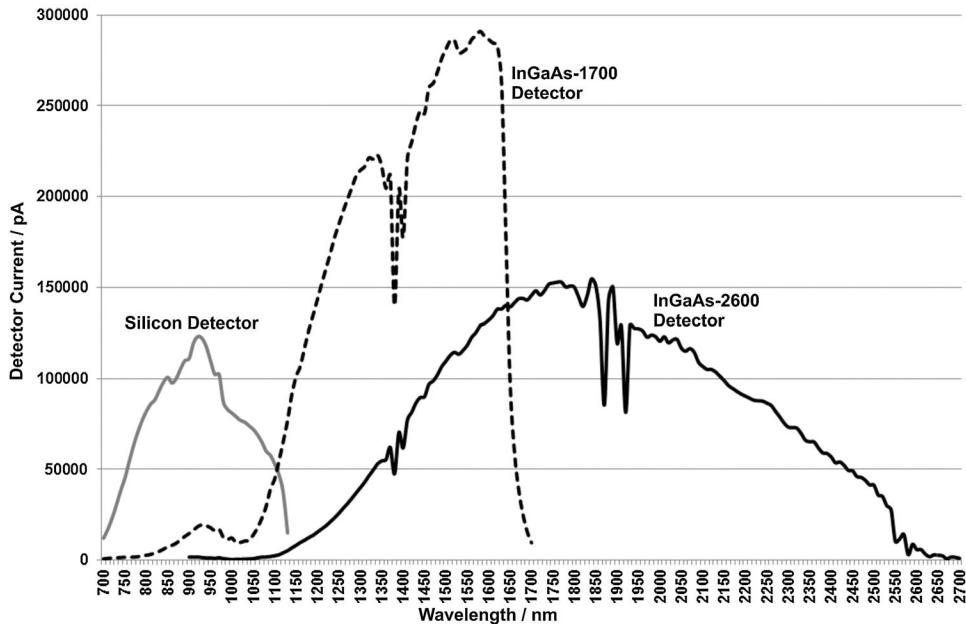


Figure 6.18 Example of detector change-over in the NIR.

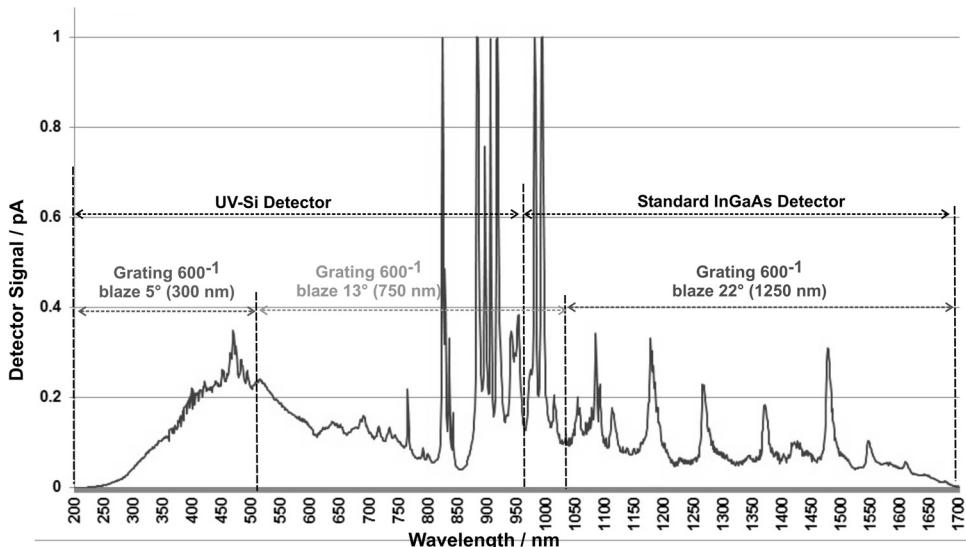


Figure 6.19 Several combined measurements of xenon in the range of 200–1700 nm.

6.6 End Result of an Illumination Monochromator System

The light flux Φ of common sources is in the range of $<1 \mu\text{W}/\text{nm}$ to $\sim 1 \text{ W}/\text{nm}$. A lamp-monochromator system is also called a “variable light source.” The decision-making point for the total efficiency of such a system is the

reproduction of the primary source—the lamp filament, the plasma arc, or the discharge. The better it is reproduced in the entrance of the spectrometer, and the higher the light-guiding factor Ω of the transfer system and spectrometer, the higher the efficiency. If only part of the primary source is re-imaged, the recovered power will drop linearly with the reproduced area. However, the power does not come only from the primary source. In practice, the source is surrounded by a halo that has reduced power emission. If the monochromator slit is larger than the re-imaged primary source, the input power distribution will be flattened, which can lead to saturation effects but no losses. In principle, it does not hurt to use a slit larger than required. Homogenous illuminated slits produce a fourfold output signal if the slit width is doubled. If the slit is not fully or homogeneously illuminated, the increase will be less. (Radiometry digs deeper into the issue.²) With regard to the output power, it makes almost no difference whether a single-stage or an additive double-stage monochromator is used. Indeed, a double spectrometer only has half the efficiency (transmission) of a single-stage version, but if two identical stages are combined, the slit width for the bandwidth required will be doubled, a reward of factor 2. Therefore, at the same bandwidth, approximately the same optical power will leave the double spectrometer, providing the additional reward of improved contrast by several orders of magnitude.

The following are realistic numbers for double spectrometers with a 1–5-mm slit width, 300-mm focal length, $\Omega = 0.04$ ($f/75$), and a bandwidth between 1–10 nm: The measured output power with a 30-W D₂ lamp will be below 1 μW . With a 200-W D₂ lamp, it will be between 1–10 μW . With a 75-W xenon lamp, a two- to three-digit μW value is realistic. A 450-W Xe will produce output powers in the lower mW range. With a 150-W halogen lamp, two-digit μW values can be expected. With a 500-W halogen lamp, it will emit in the upper-three-digit μW range.

6.7 Light Transfer and Coupling by Fiber Optics

6.7.1 Fiber guides, light-wave guides, and fiber optics

As shown in Fig. 6.20, the transmission rail of a fiber guide is the core (marked light red). The entering light is reflected at the interface layer (red) and mainly stays in the core until it leaves the system. To create an efficient interface, a thin cladding (blue) is wrapped around the core. A reinforcement material (wire netting, foil, or folded thin material, marked yellow) protects the construction and might also limit the minimum bending radius. The whole is protected again by an outer shell (most often a PVC envelope, marked black). The angles at the fiber play a key role and are a result of the materials selected. Assuming that the core consists of glass with an average refractive index (RI) of 1.5, the light entering will be treated by the general law of refraction and will travel under a shallow angle in the core. As soon as it hits the interface, it is reflected. The angle of reflection is determined by the ratio

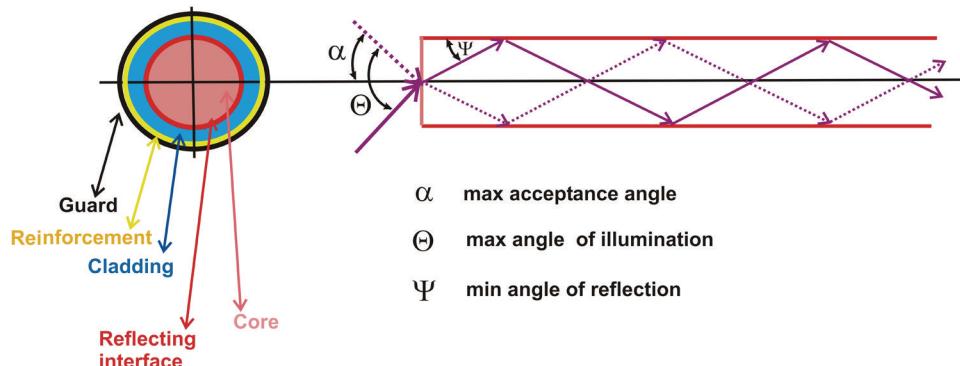


Figure 6.20 The components of an optical fiber guide, and its angles.

of the RIs of the materials that compose the interface. The combined effects define the maximum acceptance angle of entering light, which is also the maximum emission angle.

The calculation is performed by the following fiber optics equation:

$$n_0 \times \sin \alpha = (n_2^2 - n_1^2)^{1/2}, \quad (6.6)$$

where n_0 is the refractive index of the external medium (most likely air = 1), α is the maximum accepted angle of entry light with regard to the center axis of the core, n_1 is the refractive index of the core material, and n_2 is the refractive index of the cladding material.

Naturally, the RI of all materials is wavelength dependent, producing a variation of angles over the spectral range. If, for example, $n_1 = 1.5$ and $n_2 = 1.6$, the product $n_0 \times \sin \alpha = 0.4743$, resulting in $\alpha = 28.3$ deg and acceptance angle $\Theta = 56.6$ deg. Inside the fiber guide, the rays will travel under shallower angles. The internal reflecting angle ψ in the example will result from $\arcsin 0.316$, or $\psi = 18.4$ deg.

There is a large selection of materials available for both the core and cladding. They include glasses, quartzes, and plastics. The combination chosen not only defines the spectral range but also the acceptance and emission angles as a function of wavelength. Furthermore, liquid light guides and hollow-chamber systems are available on the market, extending applications into the IR and deep UV. The possible solutions are manifold; however, this book will discuss the systems for the UV–Vis–IR range, which is useful for optical spectroscopy.

6.7.2 Fiber optics for the UV–Vis–NIR range

In Fig. 6.21, the green curve (Std. Glass) shows a fiber with a glass core material, which is best for 400–2000 nm. The blue (Std. Quartz) and the pink (UV Quartz) curves are quartz cores: The latter is useful into the deep UV. However, due to the water bound in the material, strong absorption appears

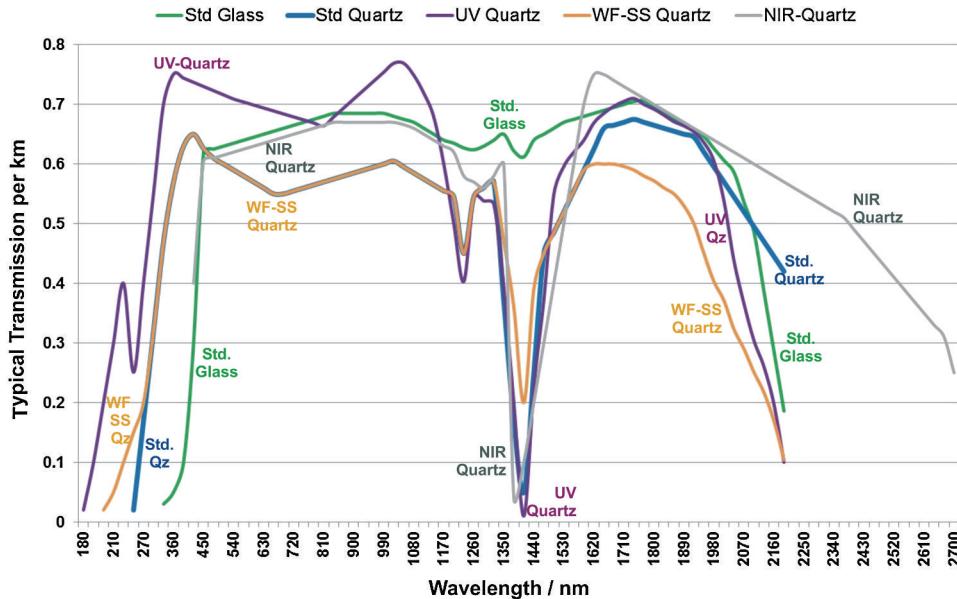


Figure 6.21 Typical transmission curves of fiber materials used in optical spectroscopy.

between 1100–1500 nm, and the former does not suffer much, though it is limited to 270 nm at the short end. The yellow curve (WF-SS Quartz) is a polarization-stabilized and water-reduced quartz. The grey curve (NIR Quartz) is suitable for the NIR range, starting at \sim 350 nm. The selected materials represent the core material and thus define the wavelength range. The curves are for general information only. The transmission values represent a length of 1 km; the coupling losses (\sim 5% at both ends) are not taken into account. Fiber cables have additional losses due to the fill factor that are also not considered. There are fiber optics that can reach deeper into the UV at the expense of the red end. Others work better into the IR, and specialty versions are only suited for rather narrow wavelength bands.

In Fig. 6.22, the dotted curve (Std. Glass) shows a fiber with core material of glass reaching up to \sim 2 μ m. The dashed (NIR Quartz) curve is a special water-free quartz core that is useful to almost 2.8 μ m. (Both are also present in Fig. 6.21.) The solid grey curve (Fl Glass) represents a fluoridated glass that reaches beyond the NIR; however, care must be taken because it can carry luminescence effects. The solid black curve (CIR) is a chalcogenide-doped material good for 1.5–6 mm. The best wide-range IR material is shown in the grey dashed-dotted curve (PIR), representing a polycrystalline material that covers 4–18 μ m.

6.7.3 Fiber optic parameters and effects

1. **Absorption:** Some of the curves in Figs. 6.21 and 6.22 show three strong transmission gaps at 1100/1200 nm, at 1300/1400 nm, and between 1800–2000 nm. These gaps are due to the water bound in normal quartz

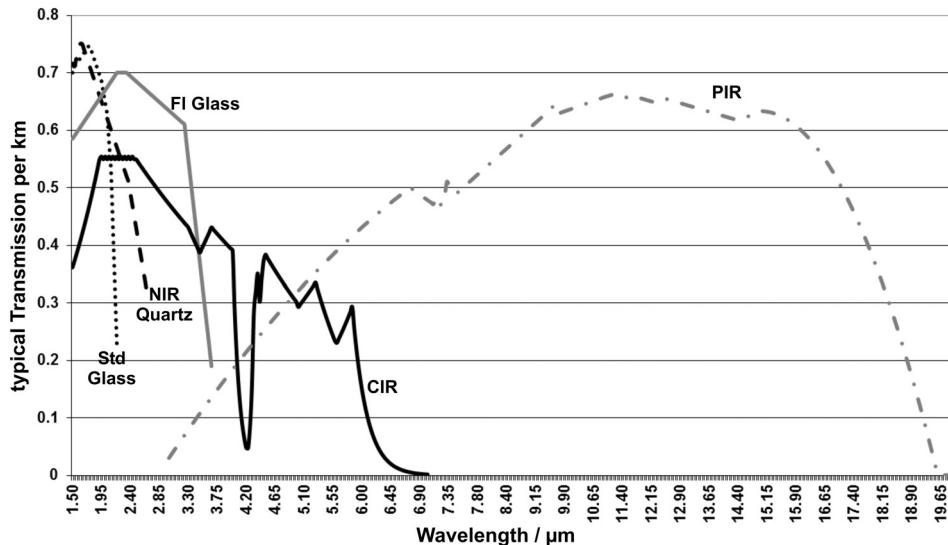


Figure 6.22 Typical transmission curves of materials used between 1.5–20 μm .

(in glass, the water is much less concentrated). Additional absorption phenomena can occur for all types.

2. **Solarization:** A displeasing effect of UV-transmitting fiber optics, solarization involves UV radiation modifying the crystalline and the molecular structure of the material, which changes the RI ratio in the interface layer and thus the transmission, as well. The local modification extends a certain distance from the entrance (depth) of the fiber and is an asymptotic effect. Thus, after a certain amount of UV energy has been introduced, the modification reaches saturation, a process called aging. The yellow curve “WF-SS Quartz” represents a so-called solarization-stabilized fiber guide. It consists of materials that are less susceptible to UV energies. To deliver a stable product, manufacturers apply pre-aging during manufacture. The material shown in Fig. 6.21 contains less water. If fibers are exposed to strong light below 230 nm, they should be checked regularly for their transmission behavior to avoid unfortunate surprises.
3. **Bending radius:** For mechanical and optical reasons, fiber guides may not be buckled too much. They either break, or change transmission substantially. The minimum bending radius results from the cross section of the core, and that of the whole product. The data sheets shall contain the curvature allowed. Many products contain a shield, which limits bending mechanically. It is a matter of fact that any cranking will change the transmission versus wavelength, even though it might be only slightly. For quantitative applications, it is recommended to fix the position of the fiber guide for as long as a series of measurement occur.

4. **Losses of in/out coupling:** As in any other optical system, fiber optics are affected by losses. The reasons are input/output reflections and diffraction at both ends. If the illumination angles do not fulfill the requirements, more wastage will occur. Like other optics, fiber guides can be coated, with the known pros and cons. If very high intensities are introduced, thermal impact might also occur and harm the functionality—permanently, if given enough time.
5. **Modes and polarization:** Figures 6.21 and 6.22 display the behavior of broadband multi-mode fibers (MMF), which are by far the most used in general spectroscopy. This configuration means that all rays within the acceptance angle are collected and transferred with almost equal efficiency. MMF behaves in a depolarizing manner: The longer the fiber is, the less polarized and the better homogenized the exiting light. Mono-mode fibers are available for short wavelength intervals, which can be polarization-conserving or polarizing, but only in one plane. Furthermore, there are fibers available that are designed for a very narrow spectral range, adding filter functions. One of those is the gradient fiber, whereby the interface between the core and cladding is not a step of the RI but a gradient. With special material selection and cross-section, special functions can be designed. For this book, the wideband multi-mode fibers and cables are the most interesting and thus are discussed.
6. **Acceptance angle:** Glass fiber guides accept and emit light at cone angles of ~ 60 deg ($f/1$); those of standard quartz types are ~ 30 deg ($f/2$); and water-free, polarization-stabilized quartz guides typically provide 25 deg ($f/2.3$). It should be repeated that there are rather wide spans in acceptance angles for every type due to the wide range of selectable materials.

Although a wide acceptance (and illumination) angle is often advantageous in general optics, that might not be the case in spectroscopy. Flexible, modular spectrometers provide acceptance angles < 20 deg ($f/3$). The majority of instruments works at angles between 15 deg ($f/4$) and 8 deg ($f/7$). It is obvious that only the 25-deg fiber and the 20-deg spectrometer fit to some extent; all other combinations should be coupled by optical adaption. Naturally, that will either lead to loss of light or to image enlargement at the spectrometer slit. Because fiber systems are available that combine wide spectral ranges (such as 220–1100 nm) with illumination angles between 10–20 deg, direct coupling or optical transfer at a 1:1 ratio is provided.
7. **Fluorescence, Raman, and Brillouin effects in fibers:** According to the materials used in the core and cladding, luminescence effects can occur, for instance, in fluoride-doped fibers. It depends on the wavelength range whether the fluorescence will disturb the measurement or not. All glass and quartz materials create Raman and Brillouin spectra. Their magnitude depends on the strength of the entering light, but all light creates discrete lines in the fiber that reach the spectrometer. If and how

strong they affect the measurement depends on the application. All three effects are often forgotten—they might be frequently negligible, but they should be kept in mind.

8. **Metal-coated fibers:** If the reinforcement (marked yellow in Fig. 6.20) is performed by a metal coating and placed around each single fiber, it disables crosstalk within a fiber bundle. It also reduces sensitivity to high temperatures, limits the bending radius, and increases robustness against breaking. The hermetic coating process is applied during fabrication by adding in-line aluminum or copper freeze-melted to the lateral surface. This process adds roughly 100 μm to the fiber diameter, which increases the distance between the cores and reduces the efficiency of the fiber cables. Thus, the decision whether to metal coat the fibers or not must also consider the surface efficiency. A single metal jacket is often applied around the whole structure of a fiber optic cable to limit bending and avoid damage.

6.7.4 “Flexible optical bench,” and a precaution about its handling

In summary, fiber optic light transfer has been shown to offer vastly increased flexibility in the configuration of the light path. However, it still behaves as an optical bench. Consequently, during a series of measurements or cases requiring intensity relations or calibration of intensity, the set-up and position of the whole system, including fiber optics, should not move. Any change will modify the described parameters and thus create a change in transfer. Sample/reference measurements will become erroneous after changes to the optical bench.

6.7.5 Typical kinds and variations of single fibers and fiber cables

In optical spectroscopy, the useful fiber cross-section starts at $\sim 20 \mu\text{m}$. The upper limit reaches into the mm range. There are even applications for whole material of up to 100 mm \varnothing . Light guides with a $\sim 200\text{-}\mu\text{m}$ core diameter probably offer the best compromise in efficiency, bending radius, and spectral range. The cladding is usually $\sim 10\text{-}\mu\text{m}$ thick. Thus, the average center–center distance of the fibers inside a cable is 220 μm . In gradient and metal-coated fibers, the cladding will be thicker. All light is transported through the core. Arriving photons that hit the cladding or the filling material between the fibers are lost. Reducing the fiber cross-section also reduces the ratio between the used and dead surface area. When the cross-section is increased, the bending radius suffers and the danger of breakage increases, in addition to the optical deficits. Further descriptions and calculations will use a 200- μm core and 220- μm fiber diameter.

6.7.5.1 Basic versions

To illuminate small and compact spectrographs with a fixed wavelength adjustment using either array or CCD detection, the most-used version is a single fiber with a 100–300- μm core. In most cases, the end of the fiber cable is

an SMA connector. These cables are also compatible with flexible, modular spectrometers, but because they provide a very small active area, a cable with a number of single fibers will be much more efficient. (The coupling is described in Section 6.8.) Figure 6.23 illustrates some frequently used versions of fiber cables. By far, the most used one is the base model shown in image 1, with or without the arm “R.” It illuminates a spectrometer system with an active slit height of 4 mm. Because the entrance slit is probably rather narrow, the single fibers are arranged as a vertical line (labeled S). At a central fiber distance of $220\text{ }\mu\text{m}$, 18 fibers are required to fill a height of 4 mm. Arranged in the shape of a rosace (labeled L), either 7, 13, or 19 fibers can be placed symmetrically. If the L side were rectangular, either 4×4 or 4×5 could be chosen. The latter is used to calculate the area efficiency of the cable: 4×5 fibers occupy $0.88 \times 1.1\text{ mm}$, or 0.968 mm^2 . The sum of the cores of the 20 fibers is an area of 0.628 mm^2 , which is close to 65%. That number does not depend much on the shape. The loss in area efficiency is one of the clear disadvantages of fiber optic cables. After that, on the S side, no further loss will occur if the slit is at least as wide as the core of the fibers.

Further descriptions of the designs shown in Fig. 6.23:

1. A fiber cable with cross-section conversion and reference arm. If, for instance, a part of the light is used to illuminate a reference detector (or a second spectrometer), one or more of the fibers can be deviated into a second arm. In the case of one fiber only, the center is usually chosen. If the cable is used at the monochromator exit, in turn, the majority of fibers will receive the light from the monochromator, while the single one might receive light from a laser or a calibration lamp, and the mixed and summed signal will reach the receiver.

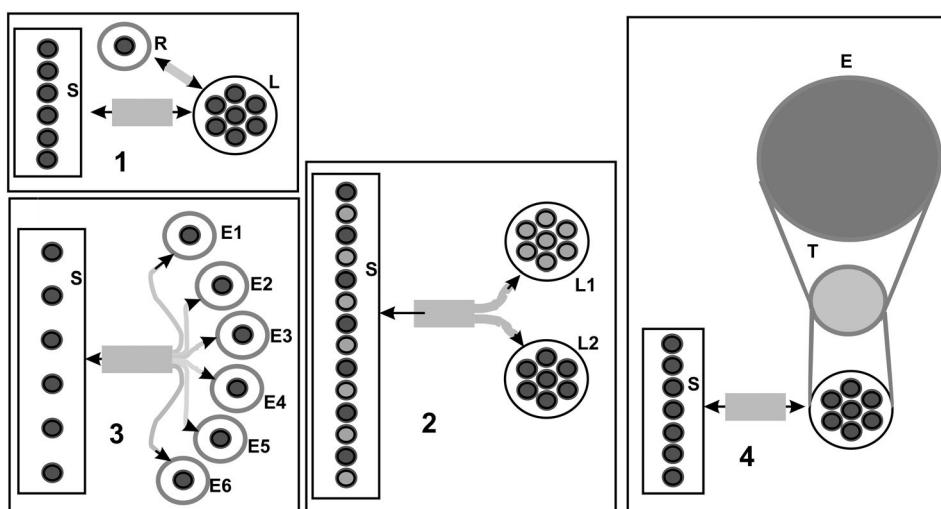


Figure 6.23 Popular fiber optic cable designs.

2. The classic case of mixing two signals equally if two light sources will illuminate the spectrometer input with the same weighting. The fibers of the two arms alternate at the S side, which will produce rather good integration and uniform transfer through the spectrometer. If a filament lamp is considered for L1 and a deuterium lamp for L2, the circular shape of L2 would be perfect, whereas the reception of light from a filament would be better suited to a rectangular arrangement. The two arms can be of different materials to fit the different spectral ranges and can have different lengths before they meet at the common place. If such a cable is used for sample illumination, the uniformity in the outputs will be rather good.
3. Independent experiments (six in the example) illuminate one imaging spectrograph with a 2D detector. The signals will reach the detector in separate vertical tracks. The fiber cable splits into six separate arms that can be of different lengths and materials (E1–E6). On the S side, the distance between the six fiber cores must be wide enough to allow separation of all tracks at the detector. To help, the read-out pattern of the detector can be adjusted in a way to avoid crosstalk. It is not necessary to use single fibers in the separate arms; small groups of fibers can also work. If, for example, a camera with a height of 12.5 mm is used, each track would be 1.8-mm high. Assuming an un-illuminated safety distance between tracks of 0.5 mm, there would still be 1.3 mm, or 6 fibers, per track. A configuration such as this, however, is not very useful at the exit of a monochromator. For the sake of homogeneity, at least the fibers at the spectrometer end need to be as close together as possible.
4. A tapered fiber cable is used here. If the spot of a light source cannot be refocused as small as the experimental end of a fiber would require, a reduction of the cross-section can be implemented in one of two ways. Shown is a version with an external taper combined with an example: If an integrating sphere has an exit of 10 mm \varnothing , but the fiber cable only has 4 mm \varnothing , a taper can be used. The calculation of the length of a taper equals the (ratio entrance/exit) \times 3. In this case, one side would have a 7.5-mm diameter, and the end would be 15.3 mm. The joint between the taper and cable can be fixed or removable; an oil with the same RI as the core material will prevent reflection losses at the interface.

One problem remains: ~35% area loss at the fiber cable surface. Alternatively, the fibers of a cable can be tapered, which is suitable for short cables (<3 m). In the example shown here, the fiber cores begin with a diameter of 500 μm before shrinking by a factor of 2.5 to reach the final value of 200 μm . Production can then proceed as normal. If the fibers are put as close together as possible at the large end, the distance of the core centers will be 520 μm . This arrangement produces an improved area yield of ~73%, an improvement of 12% compared the 200- μm fiber. A fiber cable like that will have a very limited mechanical flexibility.

Returning to the integrating sphere example, a taper like this will have the advantage of collecting all rays within the acceptance angle of the fiber material, and no internal losses will occur. Collecting the light from the sphere by classical optics and guiding it would be much less efficient.

6.8 Transfer Systems

All optical transfer systems depend on the general rule of light transfer, as explained in Section 1.6, Eq. (1.11): $M = (f_2/W_2)/(f_1/W_1)$. The magnification factor M is the product of the focal lengths divided by the sizes of objects in the two arms of a transfer system. That relationship should be kept in mind whenever different focal lengths are considered.

6.8.1 Coupling by bare optical fibers

In Fig. 6.24, the red dotted lines illustrate the emission (or acceptance) angle of the fiber, and the red areas show the theoretically illuminated area at mirror M. The green dotted lines represent the spectrometer's acceptance (or emission) angle, and the green areas represent the available mirror area. A beam stop (BS) is found in carefully designed spectrometers. It can be shaped as a rectangular aperture behind the entrance slit, which will be large enough to block the rays at wider angles without stopping beams within the acceptance angle, thus reducing stray light. In more expensive instruments, the beam stops are adjustable. If a fiber guide is mounted to a spectrometer exit, the relation is much easier. The light will be used so long as the emission angle of the spectrometer is narrower than the acceptance angle of the fiber guide. Indeed, parallel rays should be avoided to ensure reflection inside the fibers near the entrance.

A more detailed description of Fig. 6.24 follows:

1. An $f/2$ fiber guide illuminates an $f/4$ spectrometer. The rays that leave the fiber with angles up to 30 deg will be lost, to a large extent. Most of the wide angle rays that pass through the entrance slit S will be stopped by the BS. Under circumstances like this, up to 75% of the light will be lost. If light is available in excess, there might be no problems, but disturbed spectra can result due to the stray light that is created. A further drawback

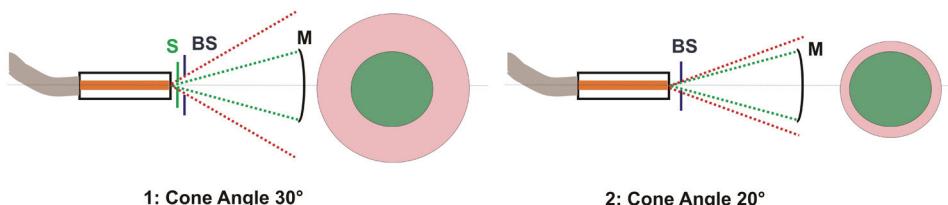


Figure 6.24 Spectrometer coupling by fiber optics without further components. S is the slit, BS is the beam stop, and M is the collimating mirror of the spectrometer.

is that it is not possible to put the end of the fiber perfectly into the focal plane, where the slit jaws are. A typical distance of 1–3 mm can remain. The focus displacement will reduce the spectroscopic performance and resolution.

2. If it is possible to select the fiber material so that it emits under $f/3$ (20 deg), only 20% of the rays will be lost. f -number ratios between 0.75–1.25 offer the choice to run the spectrometer without an entrance slit (as shown) because the fiber end will simulate the slit aperture. To achieve that scenario, the fiber end must sit perfectly in the slit plane, and the rays must travel in the perfect direction. The advantages are the high efficiency of coupling and focusing, but the cost is the missing variation of flux and bandwidth due to the fixed “slit.” This technique is often found in compact, integrated spectrograph systems designed for a certain single-fiber cross-section.

6.8.2 Coupling by lens systems

Ideal image reproduction can be realized by external optics, like the lens objectives shown in Fig. 6.25.

3. With one or more lenses, a satisfying image reproduction can be realized if the focal shift resulting from the variation of RI over wavelength stays acceptable. If, for example, the wavelength range is 400–800 nm and uncorrected quartz lenses (L1 and L2) are used, there will be a deviation from perfect parallel light after L1 within 0.27 deg, which is not a

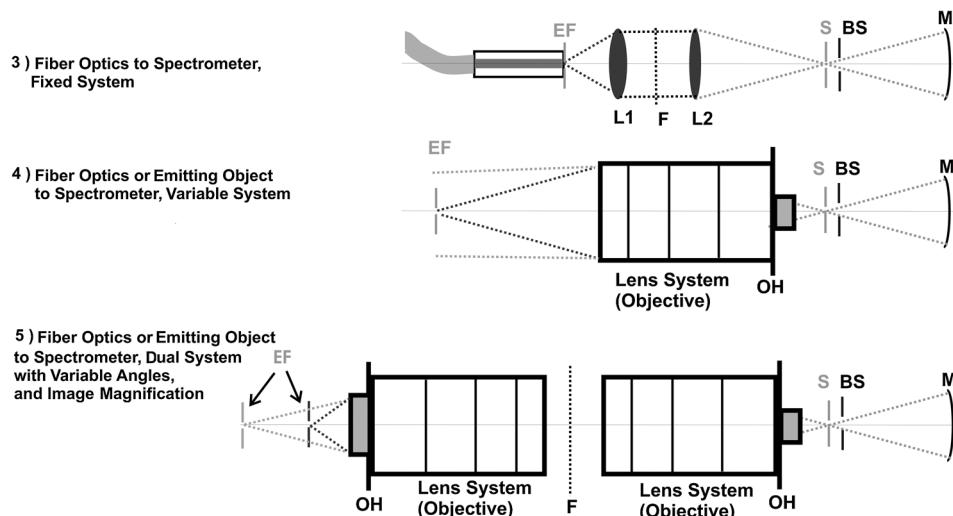


Figure 6.25 Examples of coupling spectrometer and fiber optics by lens systems. S is the slit, BS is the beam stop, M is the collimating mirror of the spectrometer, EF is the external focus position, F is the optional position in the collimated beam, L is the lens, and OH is the holder for the lens system.

problem. If L2 has $f = 100$ mm, the focal shift will be within ± 2 mm, which, in turn, will result in a focus enlargement of $20 \mu\text{m}$. It then depends on the fiber size used and the working slit width whether that arrangement is acceptable. In any case, the focus will not be worse than in case 1. Furthermore, it is evident that the two focal lengths of the entrance (i.e., 50 mm) and exit (i.e., 100 mm) of the lens system will lead to an image enlargement by a factor of 2. Working with $100\text{-}\mu\text{m}$ fibers will reproduce them in the slit plane with a size of $200 \mu\text{m}$. If so, the additional $20 \mu\text{m}$ play no practical role. Between the lenses is a space of parallel rays that is useful for optional optics such as filters. The losses at the lenses will be in the range of 15%, which is less than the losses calculated for solutions 1 and 2. The EF indicates the external focal plane, where the fiber guide ends. It can be fixed by a mechanical stop to ensure the reproducible position of the fiber cable. At the EF position, any object can be re-imaged with the required angles and be “threaded” into the measured beam. Solution 3 is equally useful for collecting light after the spectrometer.

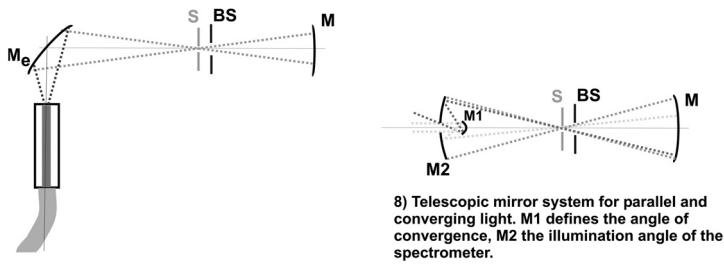
4. So long as the wavelength range is between 380–1300 nm, commercially available corrected objectives (lens systems) are suitable. A mounting plate designed for the lens system used and marked OH in the graph ensures the centric position of the lens exit at the correct distance. Like a photographic camera, the focus in the slit can reproduce an object at a certain distance, as well as parallel rays (infinite). It also works with fiber optics if the lens system is capable of the short distance that the fiber emission can require. How close the object can approach the lens system depends on the “macro capability” of the objective. The closer the f -number of the lens system fits that of the spectrometer, the better the focus and the better the efficiency. The f -number describes the ratio of the distance to the diameter of the exit lens of the objective. The distance is that between lens and entrance slit, which must be applied within the system mechanics. Until the year 2000, a Japanese camera and lens company offered UV objectives that worked down to ~ 270 nm. Unfortunately, these products became obsolete. Help might come from different specialized companies, which offer their services online, to calculate and manufacture transfer systems for the full UV-Vis-NIR range upon request. Solution 4 is useful at a monochromator exit.
5. If a small, point-shaped light source with a wide emission cone needs to be collected efficiently (with such excited experiments as Raman, fluorescence, etc.), a tricky application with two inverted objectives can help. The first lens system is inverted (its exit lens works as an entrance), and it is placed in a focal position to the light source. The rays will exit in a parallel shape, assuming that the distance to the source object was perfectly adjusted. If the object was too far away, the setting will be a finite value. If the lens-object distance was too short, it will not work. A second lens

system, which can be the same model or one with a longer focal length, is mounted “normal,” as shown in image 4. It will be set to infinite, and the diameter of its lenses will be the same as those of the counterpart. Between the systems, the rays will travel in a parallel fashion, allowing for optional hardware such as filters. The output aperture of the second system will sit so that it meets the aperture of the spectrometer. If the focal length of the first objective is shorter than the second one, image enlargement will occur in the spectrometer entrance slit. If it is the other way around, image compression will occur. Theoretically, solution 5 can also be used at the exit of a monochromator.

6.8.3 Coupling by mirror systems

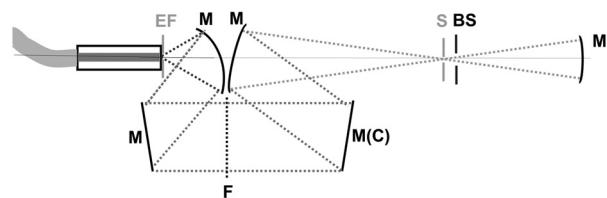
Mirror optics have the advantage of working with a constant focus over the entire wavelength range; however, they require “off-axis” light transfer (except for case 8). The solutions (Fig. 6.26) use up to four mirrors.

6. This solution has only one external mirror. The fiber guide or the object will be reproduced in the slit. Due to the rather large angles, the image will be deformed. For pure spectroscopy, this behavior will not create



6) Fiber optics or emitter with 30° cone angle, spectrometer has 15° acceptance angle, fit and transfer by a single elliptical mirror. Works in both directions.

8) Telescopic mirror system for parallel and converging light. M1 defines the angle of convergence, M2 the illumination angle of the spectrometer.



7) Fiber guide or object emission, with cone angle different from that of the spectrometer. Symmetric mirror setup comprises filter position (F) and optional imaging correction (MC). Works in both directions.

Figure 6.26 Examples of using mirrors for the coupling of spectrometer and fiber guide. S is the slit, BS is the beam stop, M is the collimating mirror of the spectrometer, M(C) is the corrected mirror for imaging spectroscopy, EF is the external focus position, and F is the optional position in the collimated beam.

problems, but it will for imaging spectroscopy. The light losses are in the range of 10%, which is less than in all other cases.

7. Very good imaging quality is provided by this version, which uses four mirrors under narrow angles. Either the fiber guide or the object is re-imaged in the slit plane. Between the two plane mirrors, additional optics (filters) can be placed. If imaging spectroscopy is the intended purpose and the spectrometer itself is not corrected, an external imaging correction can be implemented, as described in Section 4.1.6.1, Fig. 4.9. The external correction method has the advantage that the spectrometer is optimized for its major task: to produce good spectra while the external optics take care of the ray travel. The light loss is in the range of 30%, and the system has a larger footprint. The EF defines the virtual, external focal plane where the fiber guide ends or the object is re-imaged.
8. Telescope mirror systems are used in a centric relation to the beam axis. The measured light must enter either in parallel fashion (infinite) or by angular deviations up to 15 deg. The light will pass the aperture in the center of mirror M2 and then hit M1, which reflects the rays within the designated angles back to M2. That mirror has the desired focal length and diameter to fit the spectrometer, which is represented by the chain S-BS-M. The efficiency is >80%, and a special coating can improve that number. Because only unfocused light can illuminate the system correctly, it is not useful for small and nearby light sources or fiber optics.

References

1. Oriel Instruments, Inc., *The Book of Photon Tools*, Stratford, CT (2000).
2. W. Neumann, *Applications of Dispersive Optical Spectroscopy Systems*, SPIE Press, Bellingham, WA (to be published in 2014).

Chapter 7

Calibration of Spectrometers

7.1 Calibration of the Axis of Dispersion, Wavelength, and Photon Energy

7.1.1 Parameters that define the angular position of a dispersion element

The position of wavelength in a grating spectrometer depends mainly on the line frequency and the working angle of the grating, as calculated by Eqs. (2.1), (2.4), and (2.5), and derivations. A prism spectrometer will output a certain wavelength as a function of the prism dispersion, which depends on the RI of the prism at the wavelength of interest, the angle of the prism, and the illumination angle, as calculated by Eq. (3.3). However, there are more parameters that affect the actual wavelength output: the internal angles of the spectrometer, which are part of the above equations, and the actual temperature, which can change the wavelength, as described by Eq. (2.31). Furthermore the output position of a wavelength can change with the illumination of the entrance and after the instrument experiences vibrations or shock. It should be noted that any modification of the beam travel inside the spectrometer can affect the calibration. It follows that the energy axis (wavelength, wavenumber, photon energy) needs to be checked frequently and calibrated sometimes.

7.1.2 Driving a grating or prism spectrometer

Historically, there have been several different methods to place the disperser at the appropriate angle.

- a. The classical drive for scanning UV-Vis-NIR spectrometers was the sine drive; IR and Raman spectrometers have also been furnished with cosecant systems.
- b. Direct rotation by micrometric tools, with numeric display by cylindrical rolling registers. This version was only intended for static measurements, not scans. Circular working systems by no means show

the actual wavelength—they require many computations and were automated only after microprocessors became available in the 1970s.

- c. Actuation through a cam disk, which was formed in accordance to the dispersion function of the grating or prism. This method allows scanning a wavelength or wavenumbers in a linear fashion. It is still used today for prism control. A cam disk is not for general use because the curvature is personalized for a certain spectrometer with a certain disperser.

7.1.2.1 Grating spectrometers with a sine-functional drive

Because the output wavelength (for a certain grating) shows a direct relation between the sine of the working angle and the wavelength, a positioning system implementing the sine function is a direct solution that theoretically does not need corrections. The coherence between the sine/cosine and the wavelength/wavenumber is discussed in Section 4.2; this section focuses on the calibration of sine drives.

The left image in Fig. 7.1 is a sine-drive model, as described in Section 4.2.1. The horizontal structure shows the lead screw, and the sine arm is shown in two positions. The vertical position, drawn blue, shows the grating position of reflection, representing 0 nm. The angled arm position (red) represents an angle of 45 deg. The right image shows the working angles of the grating, and the relative distances, in three positions: 0 deg, 50 deg, and 70 deg. It is plain to see that the increment increases towards longer wavelengths, which is identical to the behavior of the sine curve drawn from the equations mentioned earlier. The wavelength follows the sine function with a constant multiplication factor, which represents the line frequency of the actual grating.

Sine and wavelength have a constant, fixed relationship, as shown in Fig. 7.2. Both intersect at 0 deg, which is also λ zero. Any required adjustments to the spectrometer (slits, mirrors) needs to be performed before calibration because they influence the wavelength position in the output. If the system will be used as a spectrograph, the first calibration needs to be done with an output slit anyway, to ensure proper position and resolution. The

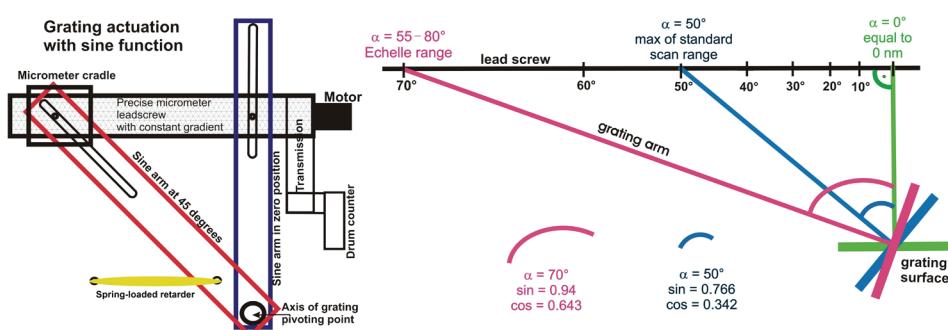


Figure 7.1 A classical sine drive, and the behavior of the sine function.

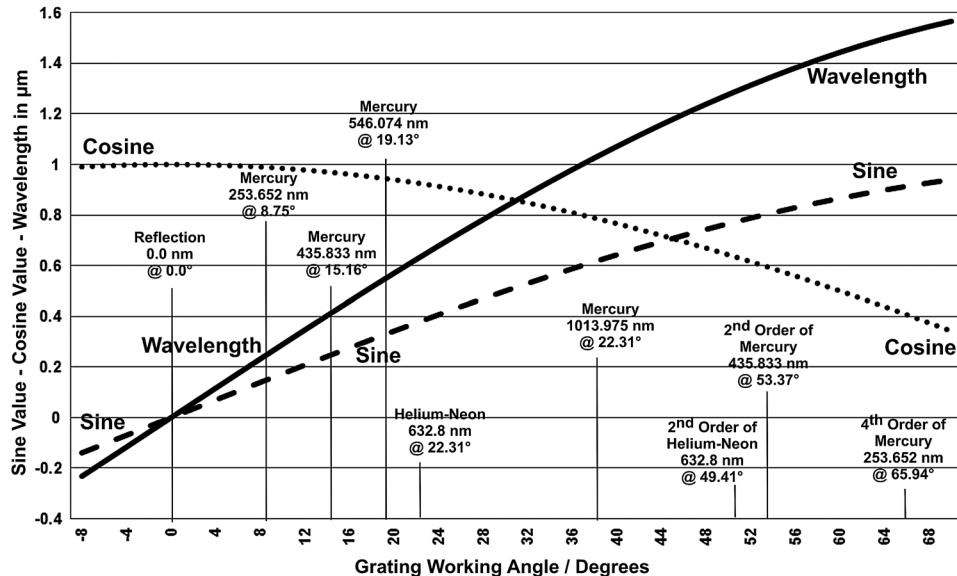


Figure 7.2 Relation between the working angle of a grating with 1200 mm^{-1} , sine, and cosine.

curves represent the angular behavior; some prominent angular calibration points are added.

7.1.2.2 Calibrating a scanning system with a sine drive

Consider the following popular example that demonstrates the sequence of the first calibration: The lead screw has a length of 200 mm, with a useful range of 180 mm. By design, the length of the particular sine arm covers -8 deg ($\sin = -0.1392$) through 70 deg ($\sin = 0.9397$). That 78-deg range is calculated so that the drum counter directly displays the wavelength for a 1200-mm^{-1} grating, resulting in a wavelength range of -230 nm to 1570 nm . That 1800-nm range will be covered by the length of 18 cm and will need 1800 turns of the lead screw. Consequently, the lead screw's precision thread has a pitch of $100\text{ }\mu\text{m}$. The system will work manually or be controlled by a stepper motor system. If that setup needs 10^4 steps per turn, it will need 18×10^6 steps for the full range and provide a precision of 0.1 pm per step. If other gratings are used, there will be a straight multiplier between turns/steps and wavelength position, which is easy to handle.

7.1.2.3 First calibration of a sine-driven system

The zeroth intersection is not a wavelength but a very-well-suited calibration point, which is also called the zeroth order. If the driving system allows scanning “through zero,” that option should definitely be used. The zeroth position at the lead screw system needs to be a perfect right angle between the screw and the sine arm. At the zeroth wavelength, all kinds of light reach the

exit slit (or the center of the output plane) and will be detected. It makes no difference whether or not the dispersion rules are not fulfilled; reflection is also marked sharply.

After the right angle between the screw and arm is adjusted, the grating is adjusted in its holder accordingly, and the drum counter and stepper system are both set to zero. The length of the sine arm needs to be set during the first calibration process (normally done in the factory). A calibration wavelength close to the longest projected working wavelength is used. Its measured position is compared with the drum counter or the theoretical number of steps other than zero. The lead screw is then slightly moved parallel in its fixture, which defines the arm length of the hypotenuse until the actual wavelength and display/steps correspond. The zero position is then double-checked. If the screw was really moved parallel, there will be no shift. Many mercury lines are available, useful in several spectral orders, to achieve that movement. Because lasers and other calibration lamps are also useful, there is always a solution to the calibration requirements.

The lead screw is then fixed next to the reference point (0 nm), probably permanently. Other calibration lines of the source will be used to check the linearity of the system. Deviations can depend on a slight angular shift of the lead screw in relation to the driving arm, which will be corrected by small moves of the end of the lead screw while the reference point is kept fixed. In short, a sine-drive-actuated grating spectrometer only needs two calibration points; other points are used to check linearity.

7.1.2.4 Parameters that can degrade the linearity

Even at zero angle, the angle between the arm and screw might not be 90 deg. Both calibration points can show the correct wavelength, but a bellow or a bay can occur in between because the length of the sine arm will travel either longer or shorter than expected. A modern microprocessor controller can take that into account and create a polynomial calibration curve; the drum counter will present the deviation regardless. If the pitch of the lead screw varies, the true spectral position will deviate locally, which cannot be corrected even by software. If the lead screw itself does not run perfectly circular, the actual wavelength will slightly vary around the true value (also difficult to correct).

7.1.2.5 Timing calibration checks and recalibration

The calibration needs to be checked after each transportation or if the system was exposed to vibrations or shocks. If the thermal environment changes more than ~ 5 °C, the calibration can change remarkably. The more coarse the line frequency is, the stronger the shift. A manual change of the grating requires a new calibration. Stepper motor systems require a further check if the system went out of control.

In most cases of de-calibration, a shift of the offset at zero λ will be the reason. A new alignment of zero will be the solution because the ascent will

not change. With a stepper motor system, all deviations within 10 turns (1 nm with 1200 mm^{-1}) can be corrected via software by changing the offset and the ascent factor.

7.1.2.6 Recalibration requirements

Recalibration calls for an appropriate light source that provides “spectral lines” in the designated spectral range. Equipment and wavelength tables are available from several suppliers;⁴ the most popular are He-Ne lasers and the Hg line source. A single-point detector at the output slit is the proper device, even if the system is intended to work with a 2D detector. It is not required to illuminate the collimating mirror and the grating fully, but it is advisable that the entrance beam runs well at the center axis of the beam travel.

1. **Zero or offset adjustment:** First, a scan through zero is performed, and the peak wavelength is recorded. Next, the sine drive is either adjusted manually or the stepper motor system is shifted so that the peak appears at 0 nm. If, for whatever reason, zero cannot be accessed, a spectral line with a wavelength as short as possible is selected (with a Hg source, probably the 253.65-nm line). The offset is then adjusted to that wavelength. Doing so creates a problem because any point other than zero carries offset and slope, as well.
2. **Slope adjustment:** The second step is to select a wavelength as long as possible within the grating’s working range (or at least at the upper end of the predicted working range). A scan around the maximum wavelength will record the wavelength found. With a manual sine drive system, the arm length might need adjustment, but that will usually not be necessary. A stepper motor system will be calibrated by modifying the “slope,” the “gain,” or the “ascent,” depending on the terminology of the manufacturer. With a 1200 mm^{-1} grating and a Si detector, the max wavelength might be either the 1013.95-nm Hg line or one of the lower lines in higher orders, as shown in Fig. 7.2.

In any case, steps 1 and 2 must be counterchecked every time the system was accessed. If the offset was not done at zeroth order, an iterating process must be exercised. That is true for both manual and stepper systems.

7.1.3 Grating spectrometers with a rotary drive

After microprocessor systems and stepper motors became available at affordable prices in the 1970s/1980s, new flexibility was designed into spectrometers. Besides the digital control of sine drives, the rotary drive became popular. It offers features such as operating multiple gratings and changing them automatically in accordance to the application. Thus, it is necessary to turn a certain position at the rotating table into a known and reproducible wavelength. This section returns to the example system with a 1200-mm^{-1} grating.

Curve and wavelength do not have a constant, fixed relation. If the grating turns around its own surface or rotates off-axis, as shown in the left image of

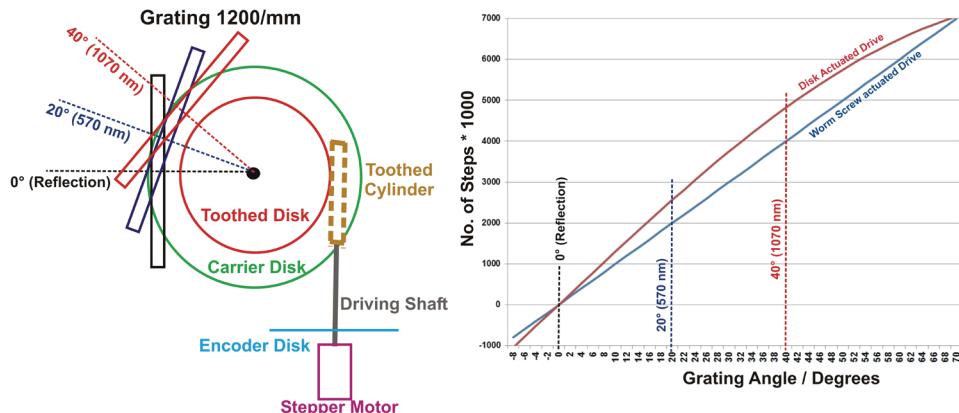


Figure 7.3 Relation between the working angle of an off-axis grating of 1200 mm^{-1} , mounted on a disk, and the angular function of a sine-driven spectrometer.

Fig. 7.3, the angular behavior does not change. It is true that the illuminated area (and thus the luminosity) varies with the working angle, but that is a different matter. The curves in the right image are marked at three positions. At zeroth order both curves—the disk-driven and the sine-driven functions—are the same. With growing wavelengths, the disk position grows faster than the sine of the angle does. Near zero, the relative difference is largest. At 7 deg, it would be 30% (not shown); at 20 deg, or 570 nm, the disk drive would still be 28%, further removed from zero than a sine-drive system with the same stepper motor. After $\sim 2/3$ of the range (40 deg in this example, or 1070 nm), the difference would still be 20%. Above that, the two curves quickly approach each other until they reach a common value again at the maximum wavelength.

The disk-driven system must follow a polynomial curve, where real-time calculations of the steps versus the wavelength take place. A small amount of nonlinearity could result in rather large errors. Therefore, the most advanced systems have integrated on-the-fly monitoring and corrections. The stepper motor can be monitored by an encoder disk, which compares the required number of steps with the actual number of steps moved. A disk system with two gratings mounted offers ~ 180 deg per grating, whereas a four-grating system would still provide ~ 90 deg per grating, sufficient for all kinds of applications.

Although a sine drive requires a right angle between the arm and lead screw at the zero intersection, a rotary drive needs a certain point of reference for zero steps. The number of actual steps from zero must be kept small; otherwise, the following operations might not produce “linear” results. Thus, the stepper system is put to zero, and the grating is adjusted in its holder. That which is created by the arm length in a sine-driven system is created by the angular difference in a disk-driven system. A calibration wavelength above the longest projected working wavelength is required next to ensure that the stepper calculations are interpolations and not extrapolations. Any outnumbering of

the maximum could lead to errors. The zero and the maximum need to be checked twice, and intermediate calibration points are required to check the linearity of the system. For gratings higher than number 1, the stepper motor system provides virtual zero positions that are readable by software. At that point, the other gratings will be adjusted to virtual zero, while the slope is calibrated like the first grating. It is essential that the upper calibration point is always higher than the longest wavelength to be measured. In short, a disk-drive-actuated grating spectrometer needs two calibration points and at least two more checkpoints to ensure linearity.

If at zero the stepper system shows an angle of ~ 1 deg or more, the linearity might suffer. If the upper calibration point is less than a wavelength used later, errors can occur. If the pitch of the stepper motor lead screw (or the pitch of the toothed disk) varies, local errors can occur. If the disk or the toothed disk does not run perfectly circular, the actual wavelength will vary and produce elliptical linearity errors; both problems are not possible to correct, and both are based on production problems.

After each transportation, the calibration needs to be checked; this is also true if the system was exposed to vibrations or shocks. The strength of the impact of environmental changes in temperature highly depends on the size and construction principles of the rotary system. Variations will occur after thermal changes regardless. A new calibration is required after adding or changing a grating. If the system went out of control, a check is required. Most cases of de-calibration are caused by a shift of the offset. A new alignment of zero is the solution because the ascent will not change. The requirements, tools, and proceedings are identical to sine-driven systems.

7.1.4 Calibration of the field output

It is assumed that the center of the field detector is at the same position as the center of the output slit. In other words, the same procedures described earlier are used for the calibration of the center wavelengths. According to Eqs. (2.1), (2.5), and (2.17)–(2.19), it has been proven that the dispersion (and thus the distances between wavelengths) slightly changes over the wavelength axis of the output field.¹

The left graph in Fig. 7.4 shows the wavelength position versus the output position when a symmetric, 500-mm spectrograph ($\epsilon = 10$ deg), equipped with a 1200-mm^{-1} grating, is set to 570 nm. The curve looks linear. For most applications, the linearity might be fine. The right graph shows the differential dispersion of the same situation. The dispersion differs in the fourth decimal. Unfortunately, it is a very complex process to put all parameters that influence the wavelength distribution in the field plane into a single equation, as the paper by Lindrum and Nickel¹ demonstrates. Broken down, it means that every other angle in the spectrometer will affect the output distribution, such as turning the grating to a different position or changing the tilt angle of the detector for the best focal behavior over the plane, as described in Section 4.1.5.

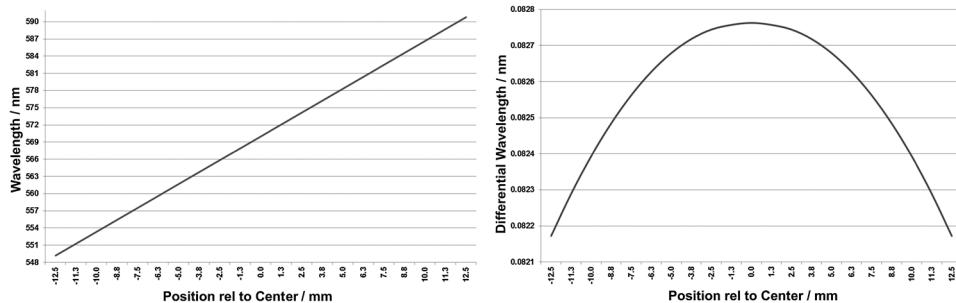


Figure 7.4 Wavelength distribution in the output plane of a spectrograph.

7.1.4.1 Output dispersion as a function of the lateral position in the field output

Figure 7.5 shows four curves that demonstrate how the local dispersion changes over the lateral (wavelength) axis of the field output. (The reasons and coherences are discussed in Section 4.1.4.) When running a camera, the precise position of wavelengths needs three calibration points within the recorded spectral interval, and a curve fit will create a calibration function. If three known signals are not available, a single point plus a constructed fit curve might provide the required precision. In any case, at least one calibration signal is required to check the calibration with the camera. Because different positions of the grating will also change the slope of the curve, moving the grating to test the calibration is only useful near a spectral vicinity within ± 2 deg of the grating's angle.

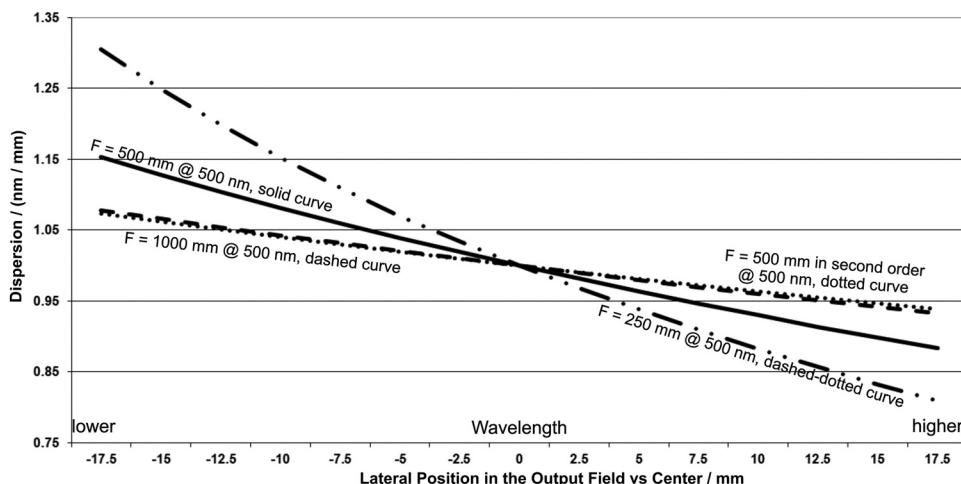


Figure 7.5 Four typical dispersion curves.

7.2 Calibrating the Axis of Intensity, Signal, and Illumination

Applications that require a calibrated intensity axis almost always deal with the portability of intensity data, thus eliminating system dependence. The common name for that kind of measurement is “radiometry.” (See this text’s companion book² for a more detailed treatment.) The radiometric nomenclature and its equations were introduced in Section 6.2; this section addresses the practical aspects of reliable intensity calibrations.

7.2.1 Requirements for a useful calibration and portability of data

To achieve reliable signal integration, the light source needs to fulfill several parameters. It must spread over the full wavelength range that is proposed for the measurements of unknown samples. The calibration source should not contain strong and steep changes of signal versus wavelength. The intensity distribution is best if it is on the same order of magnitude as the unknown samples at all wavelengths. If the calibration data are planned for publication, they need to be stored in open ASCII code and probably portable graphics. Of course, the intensity calibration demands a validated wavelength calibration.

7.2.2 Light sources for radiometric calibration

Not many types of light sources are useful for the calibration of spectrometer systems. The first choice is a filament lamp, such as tungsten and tungsten halogen lamps with a quartz window.³ Useful data can be produced between ~280 nm (in certain cases, 240 nm) and 2800 nm. The lower end is defined by the filament temperature, and the upper by the bulb material. In the range above 2 μm, thermal radiators such as blackbodies are the first choice. In the UV range, deuterium lamps serve between 120–400 nm. If very narrow spectral ranges must be calibrated, LEDs of a relatively wide bandwidth are also useful. Xenon arc lamps are not useful because the arc’s position and shape can vary over time, even if that difference does not change the output power.

If the data produced with the help of a calibration source are to be published, the light source used must be calibrated and documented by a certified calibration lab. If the calibration is only used internally and no published data depend on it, an uncertified source might be sufficient. Typically, the calibration source documentation will consist of the wavelength and light-power data pairs. The wavelength is normally listed in 1-nm steps and a 1-nm bandwidth, whereas the intensity is mainly listed in terms of the radiant power $\Phi_{(\lambda)}$ [W]; other expressions are possible. Most formats can be converted to the kind of data required by using the equations listed in Section 6.2. The documentation provided by the certified lab will also contain the period of validity of the calibration and the maximum working hours within that period. Other important parameters are the voltage and current used at calibration; sometimes even the environmental temperature is indicated. The data need to be stored in the same format that the spectrometer

system uses for later data processing. That curve will be called “OD,” for “original data.”

7.2.3 Procedures to produce reliable calibrated data

Before the calibration source is turned on, the optical configuration should be put into the final state. After acquisition of the signal from the calibration source, not even a small modification to the beam travel is allowed—that means the calibration source and the device under test (DUT = unknown source) should be placed at the same place and pass the same devices identically. If fiber optics are involved, they need to be fixed before measurement and kept in place. Even small changes to the fiber optics will change their transfer characteristics; they need to be treated as an optical bench. If the DUT cannot be placed at the calibration source, some linear calculations need to be done to fit the data. The entrance slit of the spectrometer must be fully illuminated by both the calibration source and the DUT. It would be ideal if the illumination image at the slit were homogenous. If that is the case, the slit’s width at the entrance and exit can vary during the session, but note that a squared intensity effect will take place upon changing the slit widths. If the DUT does not fully illuminate the slit height, the calibration source measurement will be limited to the same height used by the DUT experiment.

The spectrometer and detector must be sufficiently warmed up first. All parameters are either set to the data provided by the calibration source documentation or to close values; all settings will be noticeable. Constant monitoring of the source supply is required. After warming up the calibration source, the first scan is taken and stored according to its documentation. If the data look reasonable, at least two more scans are taken, and all of them are stored independently. The calibration source should not burn longer than required to save operation time. Next, the three (or more) curves are averaged, and the STD of the single curves versus the average is calculated for each data point. It must then be decided whether the STD promises data reliable enough for future measurements. For instance, if the data reduction will stay within an uncertainty of 3%, the STD needs to be within 1% to provide a good confidence level. If that is true, the average curve created by the three (or more) single calibration runs is used as reference “R.” The single curves will be filed regardless. The reference R must then be turned into the instrument response curve “RC.” The quotient of the lamp data OD (from the documentation sheet or file), divided by the reference curve R, results in the instrument’s response curve RC, which is stored. To test the procedure, the response curve RC can be multiplied by the reference curve R. The result must be the numerical lamp data from the document, OD. If that is fine, the response curve RC will be saved; it can be headed with the radiometric parameters of the document, such as nW/nm. Later, the DUT data are recorded and stored as an unknown sample “S.” The multiplication of $RC \times S$ provides the radiometric-corrected (calibrated) final sample result “SR,” representing the DUT emission that was transferred into the spectrometer during the measurement.

It might be a good idea to turn off everything, start from scratch, move the grating and slits, and bring them back to the proper settings. Repeat the R measurement as an unknown sample S and multiply $RC \times S$. The final result will be identical to the documented data of the calibration source OD, within the acceptable tolerances. The unknown samples can then be measured, and their power can be recovered. The following are the curves involved:

- **Original data:** OD [nW/nm];
- **Measured data from the calibration source:** R [μ A] (detector signal, can differ from the current);
- **Instrument response:** RC from OD/R [nW/(nm \times μ A)];
- **Unknown sample (DUT):** S [μ A] (same dimension as R);
- **Final result:** SR from $S \times RC$ [μ A \times nW/(nm \times μ A)], result is in nW/nm.

It is obvious that the RC needs to be checked occasionally. The frequency depends on the proposed STD; experience will help with the decision. In order to save operation time of the original calibration source, it might be a good idea to create a secondary standard, which can be one of the DUTs and which provides a good internal reference that can be kept safe. The final result SR, produced from the daughter reference, can be stored and used for daily comparison. As long as the data captured from it are identical, everything will be well within the acceptable tolerances—the system is fine and needs no recalibration.

In Fig. 7.6, the curve (enhanced by dots) OD is the graphical presentation of the original calibration source data. The curve R is the measured data, or the average of several measurements. The curve RC is the result of OD/R (the

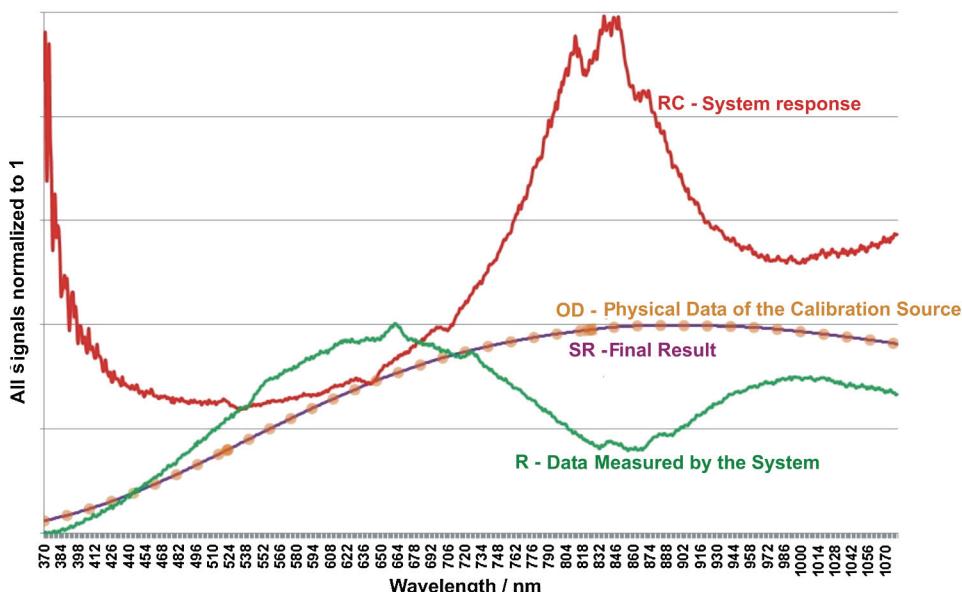


Figure 7.6 Sample curves of an intensity calibration.

system response). Finally, the curve SR presents the final result. As in the example presented earlier, all data were taken with the calibration source, and OD and SR are identical, which proves validity.

7.3 Transfer Efficiency of Spectrometers

7.3.1 General behavior

The wavelength-dependent efficiency T of a monochromator or a spectrograph is mainly influenced by the grating. For the mirrors, the reflection efficiency is generally found between 0.88–0.96, in the wavelength range of 190–2800 nm. The exact value depends on the wavelength and coating. If these are not known, 0.9 is a good guess. For gratings, it can be generally estimated that, at the optimal wavelength λ_1 , the dispersion efficiency can reach up to 0.9. The reflection losses of the front face depend on the coating and will be in the range of 3–15%, resulting in a total efficiency of ~0.8. Taking into account the efficiency of 2–4 mirrors, the total transfer efficiency T of spectrometers in the UV-Vis-NIR range will be 0.4–0.7. It is no mistake to estimate a total T of 0.5 at the blaze wavelength. Apart from the blaze wavelength (λ_1), the grating efficiency will drop towards the UV to reach 50% of the maximum (generally at about $\lambda = 0.7 \times \lambda_1$). The same value will probably be found at $\lambda = 2 \times \lambda_1$ towards the IR. In both cases, the total T is about 0.25. The interval in between is generally seen to be the optimal working range within one order. Changes in polarization and anomalies are not considered. Because all of the effects are found twice in a double-pass or double spectrometer, the optimal range will be shortened. Investigating T at one wavelength or for a certain range can be difficult. Also, it is questionable whether the efficiency of the spectrometer alone is useful; thus, it is often combined with the light source and/or detector. For instrument developers, the T behavior is a tool of optimization. If T only needs to be known for one or several wavelengths, or a narrow range, a laser can eventually be used, making the process rather easy. Because T does not depend on the aperture and is assumed to be independent from the used areas of the spectrometer, the beam can probably be sent into it without modification. For strong lasers, of course, attenuation might be required.

7.3.2 Measurement of transfer efficiency

Before measurement, all parts that are not required or that could be damaged are dismounted. The laser beam is then sent through the fully open entrance. Right after the entrance, an energy meter is used to record the intensity. The grating is then moved to the wavelength and order under test, and another measurement is taken before the beam leaves the exit. The amplitude ratio is T . Some aspects must be taken into account. If the laser has more than one emission line within the response range of the energy meter, filtering is required. The filter can be a second monochromator. Note that the exit light

will be dispersed. It must be ensured that the whole bandwidth available after the entrance is also recovered at the exit. Placing the energy meter into the DUT might require the system to run exposed, which would necessitate a dark room. The ratio of the exit-to-entrance signal will directly lead to the transfer efficiency T .

Evaluation over a wide range is more complicated due to the lack of lasers or diode sources. One of two arrangements is required: (a) a source covering the range, plus a tuning filter or set of filters, and some optics to enter the spectrometer under test with a narrow angled beam; or (b) a wideband source that feeds an illumination monochromator that covers the range of interest. The spectrometer under test will follow as a second stage. The first stage consists of a variable light source that delivers defined light to the spectrometer under test. The entrance illumination needs to be geometrically smaller than the fully open slit, and the illumination angle should not be wider than half the aperture of the spectrometer under test in order to keep the light concentrated. Because the light source already delivers dispersed light and the spectrometer under test disperses again, the wavelength of interest can cover a rather large area before the light leaves the exit (point of measurement). The energy meter in Fig. 7.7 (grey ellipse) is shown in two relevant places and must be large enough to cover all light after dispersion and a possible vertical stretch. By taking a series of ratios, T can be recorded within the required wavelength and order. Depending on the behavior of the first unit, order-sorting filters might be required. If the first monochromator is a double monochromator, it can also be used for the analysis of stray light, as demonstrated in Chapter 8.

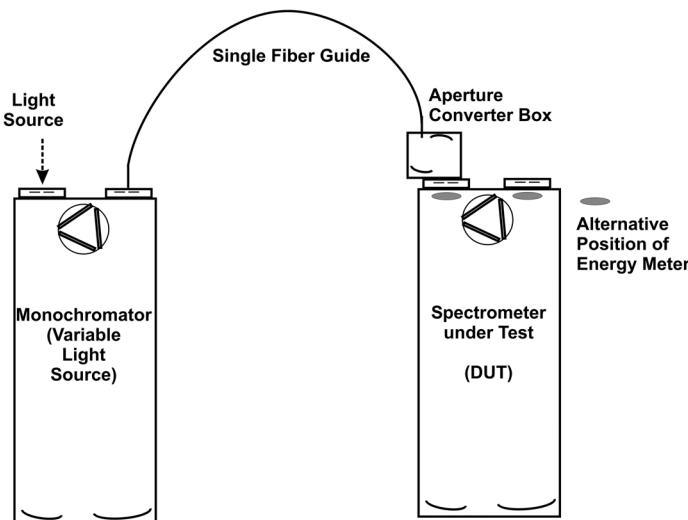


Figure 7.7 Setup for the analysis of the transmission vs. wide-range wavelengths.

References

1. M. Lindrum and B. Nickel, "Wavelength calibration of optical multichannel detectors in combination with single- and double-grating monochromators," *Appl. Spectroscopy* **43**(8), 1427–1431 (1989).
2. W. Neumann, *Applications of Dispersive Optical Spectroscopy Systems*, SPIE Press, Bellingham, WA (to be published in 2014).
3. G. F. Knoll, *Radiation Detection and Measurement*, Fourth Edition, John Wiley & Sons, New York (2010).
4. Oriel Instruments, "Typical Spectra of Oriel Instruments Spectral Calibration Lamps," Oriel Instruments, Stratford, CT, <http://assets.newport.com/web/Documents-EN/images/29590.pdf>.

Chapter 8

Stray and False Light: Origin, Impact, and Analysis

Stray light, or false light, is one of the effects in optical spectroscopy that are absolutely unwanted but (almost) unavoidable. The name “stray light” describes uncontrolled, vagabond light. It comprises scattered light, uncontrolled reflections, over-illumination, and the overlay of spectral orders. The origin of stray light and its impact on the desired signal is complex and diversified; the same is true for the prevention and reduction.¹ For example, the analysis of an LED will create rather little stray light compared to illumination with strong, broadband light. The response of the receiver, detector, or experiment at the different wavelengths (available as desired and undesired signals) is also important. Consequently, stray light must be viewed application-dependent, which is why an entire chapter is dedicated to the issue; the impact of stray light is treated separately in this book’s companion volume.²

8.1 Origin of Stray Light

The illumination of components in an optospectroscopic setup challenges the creation of stray light at the same time. No technical component is perfect. Mirrors, lenses, windows, filters, gratings, and prisms—all provide imperfect surfaces that deflect a small number of rays in unwanted directions. Inside a spectrometer, the components can be kept clean rather easily. In an open room, however, dust and dirt are not simply stopped. All kinds of residuals, roughness, or leaks are sources of stray light. The illumination of a spectrometer through a slit or aperture is critical. Diffraction might occur, or the cone of light might be wider than the acceptance angle and overfill the instrument. When calculating and adjusting the illumination, it is useful to base such actions on the reduced grating width at the highest working angle. This precaution will produce the smallest width of the optical beam and avoid

over-illumination. If no edges are hit by the light beam, one possible source of trouble is avoided.

The grating can be the strongest source of stray light inside a spectrometer because it produces scattered light; this is due to the surface structures (the diffraction lines), also called grooves. Large differences in scattering are found throughout the available types. A holographic grating without further processing will comprise almost perfect sine structures that scatter very little. Optimization methods, such as etching, modify the sine structures in favor of creating a blaze behavior for the planned working range. That modification will, in turn, add irregularities to the surface that increase the stray light level. Ruled gratings are burdened with imperfect structures in any case, mainly in the valleys, and at the crests of the line structures. The stray light, created by a ruled grating, is larger by a typical factor of 3–10, compared to a blazed holographic one. In all cases it is true that gratings with a higher line density create less scattering than those with a lower density. This behavior is primarily due to the magnitude of modulation, which decreases with increasing density. Additional coating of a grating will level out some of the irregularities and thus lower the level of stray light. Coating also reduces scattering from mirrors and lenses. Prisms show small scattering effects that are not stronger than lenses and other homogenous components. Compared to gratings, the prism is generally a much weaker source of stray light. Finally, the detector can also produce stray light or reflection effects. A detector surface that retroreflects part of the arriving light back into the spectrometer will disturb the regular beam travel. Diffusion effects in the detector's cover coating or window will, if they occur, create effects that are like stray light. Of course, a single-point detector mounted behind the exit slit will hardly create problems. Often, its window is curved (i.e., PMT bulbs) or tilted, and almost no back light can enter the spectrometer, thus avoiding the problem. Area detectors, on the other hand, are not separated by a barrier, such as the slit. The light reflected from the window's surface can return to the beam. That behavior is suppressed if the detector contains a vertically wedged window (approximately a degree in the vertical is sufficient). If that is not the case, a small vertical tilt of the housing or the mounting flange might help. In the case of line array detectors, the tilt creates no drawback; in the case of 2D detectors, it depends on the focal change over the surface. Wedged detector windows have a second advantage because they reduce interference effects.

The solid black curve in Fig. 8.1 represents the behavior of scatter, which can be the major basis of stray light. It radically rises below 350 nm. The dashed-dotted curve is the typical efficiency of a UV-Si detector (diode, array, CCD). The grey curve is the normalized beam density of a 100-W halogen lamp, while the lower dashed curve shows a 30-W deuterium lamp referenced to the halogen curve. The combination of the curves results in the fact that relative measurements (absorption, reflection, fluorescence, etc.) can be critical below 450 nm. If the detector is a PMT (dotted curve), the detection

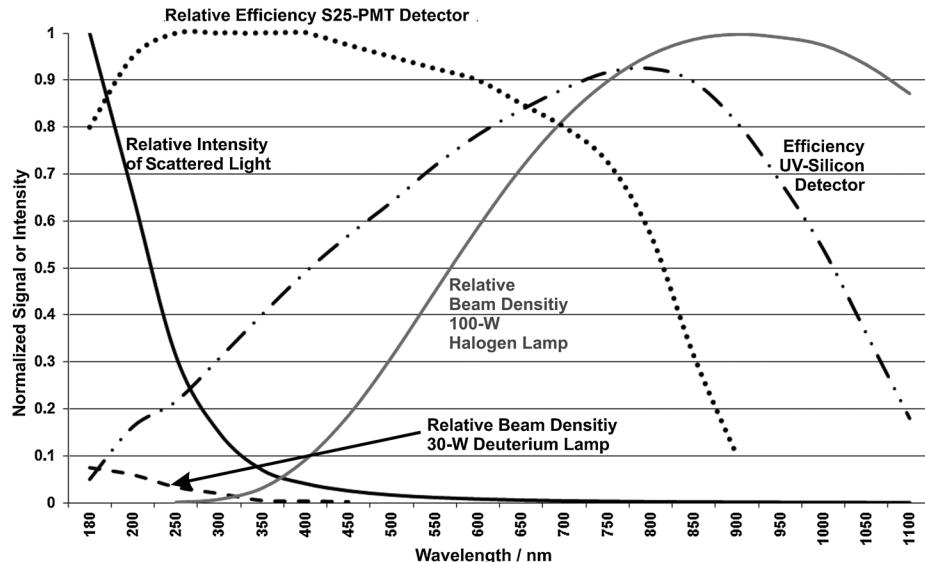


Figure 8.1 Some parameters that can affect the magnitude of disturbance.

efficiency will increase in the critical range, which will lead to an improved ratio of measured light to stray light in the UV.

8.2 Impact of Stray Light

8.2.1 Disturbance in the application of discrete spectral signals

The creation of stray light is minimal when measuring a single or a few spectral lines. If, for example, the light from an LED with a 50-nm bandwidth is analyzed, no disturbance will originate from outside of the range of interest; only the signal itself can be the source of noise. Conversely, noise can come from reflections, over-illumination, and scatter from the components. Daughter signals from the primary source can occur, which can be separated by filtering or by changing orders. The unwanted extra signal can reduce the real one by distributing energy to different places and can widen the appearance of the bandwidth of the signal. Effects like that can be analyzed by additional bandfilters. In all cases, the stray light will reduce the signal/background ratio. With monochromators of focal lengths >250 mm and gratings of 1200 mm^{-1} or more, no stray light levels above 10^{-5} are expected. In principle, the same is true for laser light with only a few lines; even high optical power can produce nonlinear effects, which in turn lead to increased stray light levels.

The analysis of collections of spectral lines, such as those from Pen Rays[®] or plasma emission, is a similar circumstance. If the detector is a parallel one in the output field, the stray light level will increase naturally, but it is hard to

quantify. It is also not easy to estimate and judge the background between the peaks because the background can come from the light source itself. There are basically only two methods for finding the source of the intermediate background: an additional measurement with a grating of higher dispersion, and a reference measurement with bandfilter (a third method involves a double spectrometer). The comparison of the results follows the different measurements. If the results are similar, the background is real and not a disturbance. If a spectrometer is configured to be a monochromator, the probability of light passing the exit slit under forbidden angles is very small. If the exit slit is removed and an area detector is mounted, the probability of rays outside the designed angular range that reaching the field and being detected increases strongly. That scenario, in combination with the far larger output area itself, increases the chance of disturbance remarkably. Experiments resulted in factors of 10–100, comparing the field output versus the slit operation. If the difference is not known, a multiplication factor of 30 can be used for estimations.

The graph in Fig. 8.2 can be used for the estimation of stray light. It assumes a spectrometer in monochromator mode that creates a stray light level of 10^{-5} between 450–600 nm. Two strong lines at 488 nm and 514.5 nm are considered, which combine to a 3500-mW optical power. Thus, 35 μW of stray light will be diverted and distributed over 450–600 nm; that means that 0.23 $\mu\text{W}/\text{nm}$ will be added to the real signals and the background. The fact that the weaker laser lines create additional stray light is omitted. The lines at 501.7 nm (nominally 300 mW) and 520.8 nm (nominally 8 mW, a very small peak in the graph) are measured and found to be enhanced by 0.23 μW , which

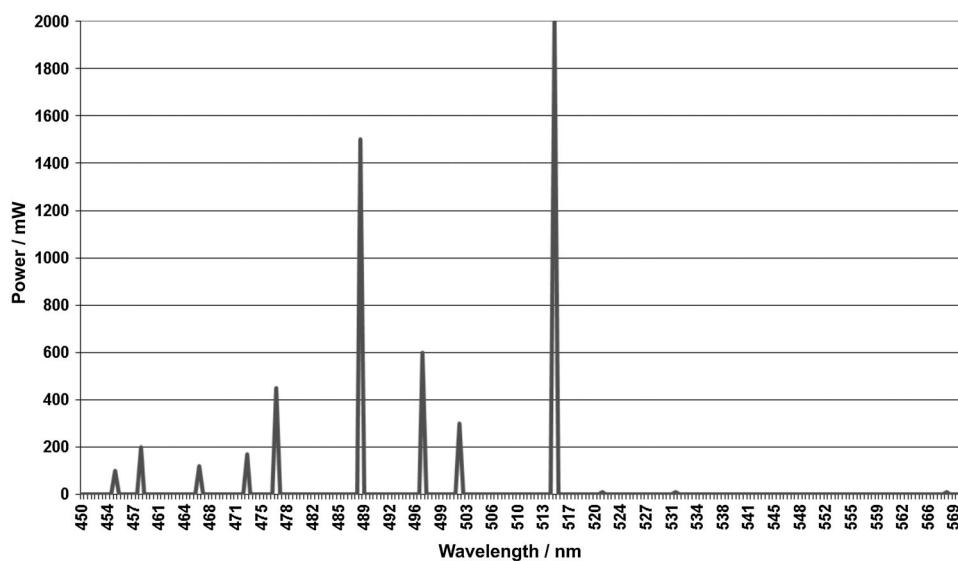


Figure 8.2 Lines of a 10-W argon-ion laser in the vicinity of 500 nm.

might be within tolerance. The same spectrometer in spectrograph mode can show a noise level of 2×10^{-3} (factor 50). It will distribute a stray light total of 1.75 mW, which makes $11.67 \mu\text{W}/\text{nm}$. That amount will add very little to the 501.7-nm line, but the 520-nm line will increase by 1.5%.

8.2.2 Disturbance in the application of broadband spectral signals

One of the most difficult challenges for a spectrometer is selecting a small bandwidth out of a strong, broadband entrance spectrum and representing it correctly. Apart from the linear effect of distributing stray light at higher levels compared to limited entrance bandwidths, matrix effects can appear. The example from the previous section is extended with a 100-W halogen lamp to analyze the remnant radiation below 300 nm. The stray light level in monochromator mode will still be 10^{-5} , with an estimated 2×10^{-3} in spectrograph mode. For simplicity, assume that the detectors reach from 200–1000 nm and that all noise signals are the same at all wavelengths (which is not true in reality). Furthermore, assume a monochromator bandwidth of 1 nm and also a pixel bandwidth of 1 nm in spectrograph mode. The transfer function of the spectrometer, mainly defined by the grating, will be hypothetically constant over the wavelength. This simulation is based on Fig. 8.1. Out of the 100 W from the lamp, the illumination system brings 0.6 W into the spectrometer, spread from the UV through 1000 nm. To find the distribution of optical power, one only needs to read the scale of Fig. 8.1. The integral of 600 mW, distributed to 200–1000 nm, gives a median value of 0.75 mW/nm. Hence, the monochromator creates a median disturbance of 10^{-5} , or 7.5 nW/nm. In spectrograph mode, the disturbance will increase to a median of $1.5 \mu\text{W}/\text{nm}$ over the range of 200–1000 nm. In monochromator mode, the stray light will create errors of <0.1% above 400 nm. But at 200 nm, where the lamp under test provides practically no radiation, a value of 7.5 nW will be presented. Even at 300 nm, at a real power of $7.7 \mu\text{W}$, ~0.1% will be added.

Applications in radiometry cannot accept errors like that, which necessitates the use of a double monochromator. That device will suppress the stray light to the range of 10^{-8} or less, reducing the errors by at least three orders of magnitude. When applying a parallel detection system, the displayed value at 200 nm will be $1.5 \mu\text{W}/\text{nm}$, and at 300 nm it will display $9.2 \mu\text{W}$, or 23% more than real. Even at 400 nm, where the real lamp power is 0.1 mW/nm, a value of 0.1015 mW/nm will be measured. Additional problems will result from overlaying the three spectral orders.

Thus far, the spectrometers and detectors have been assumed to treat all wavelengths equally. To correct that, the known result is convoluted by the real efficiency of the silicon detector. Because the light source carries 83% of its total power (in the range of interest) between 600–1000 nm, the majority of vagabond photons will come from there. Unfortunately, the Si detector responds to that range with an efficiency of ~0.8, whereas at 300 nm its

response is only 1/4 that. Hence, we need to multiply the disturbance, which occurs in the UV, by a factor of 4. That will result in an increase of the disturbance in monochromator mode to 30 nW/nm, or 0.25% at 300 nm. With parallel detection, the expected error will increase to 6 μW , which leads to a display of 13.7 μW at 300 nm (true value: 7.7 μW). Above 450 nm, the additional error slowly drops. In comparison to the Si detector, we estimate the same situation with a PMT, and find a very different picture. In contrast to the Si model, the PMT below 500 nm is far more sensitive, but above 500 nm the detection efficiency drops. Refer to Figs. 5.3 and 5.6 for details.

Spectral signals >900 nm, adding up to $\sim 1/4$ of the power within the interesting range, are not detected at all by standard PMTs because the working range is limited to ~ 850 nm. However, in the UV, the effects of the disturbance will strongly be minimized. A good measure for the impact will be 1/10, or 0.75 nW/nm, leading to the same relative decrease in the convoluted values. Now for a closer look at the grating. If it were optimized (blazed) to 250 nm, it would prefer the range of 200–300 nm by a factor of ~ 5 over the range >700 nm. The impact of the stray light would then drop strongly, and furthermore, the final system output would be much flatter over the whole range.

In conclusion, the reasoning proves that it is worthwhile to spend time and money on system optimization. If there is a real need to apply parallel detection, it might be worth trying to split the large range into three intervals that can be acquired sequentially and recombined by software. While moving the grating, an automatic order-filter change will be added, and probably (depending on the comfort of the program) a modification of the illumination time, too. Besides eliminating the spectral orders, the split into three intervals will drastically reduce the stray light if the dispersion fits the shorter wavelength intervals.

8.3 Analysis and Quantization of Stray Light in Spectrometers and Spectrophotometers

The definition and measurement of stray light should happen as close as possible to the future applications. If a spectrophotometer is purchased (or a Raman spectrometer or a fluorometer), one can expect useful and application-oriented numbers for stray light to be provided by the manufacturer. Furthermore, the analysis is not very complicated. However, a customer who buys a building-block system for more flexibility and a wider range of applications cannot expect to be given the relevant stray light numbers. Building-block spectrometers are sometimes not supplied with that data at all; if they are, they will most probably be valid for a 1200-mm^{-1} grating that is holographic and optimized to 500 nm. The stray light might be measured with a He-Ne laser at 632.8 nm. In most cases, the detection is performed with a bandwidth of 1 nm. If the detector is adjusted to full scale at the laser peak,

and the stray light is acquired at a distance of 10 bandwidths, which means at 622.8 and 642.8 nm, then the result will be $<10^{-5}$ for a 500-mm monochromator. If the application uses light from a gas laser, diode laser, or LED in the visible range, the specification is still useful, but if only a change to the parallel detection is achieved, the noise will increase remarkably, eventually by one order. If the source contains discrete lines that are at least 20 bandwidths apart, the probability of matrix effects will be small, and the noise floor might stay constant. The wider the interval of radiation grows, the closer the discrete photon energies melt together, and the faster the total stray light will increase.

Unfortunately, it is not possible to express the impact in numbers; it needs a measurement. If the wavelength of the laser were shifted from ~600 nm to the range of 300–350 nm, thus doubling the energy in eV, and if an optimized 1200-mm⁻¹ grating were mounted, the system would face a noise floor of ~15-fold. That number would greatly improve if a grating of 2400 mm⁻¹, blaze 300 nm, were used. The typical impact on the output signal would then only be a factor of 5. So long as the spectral lines are well separated, the measurement and stray light analysis is rather easy. If the experiment has spectral lines of high density, as is the case in many emission experiments or as shown in Fig. 8.2, the picture will change. If the source itself produces a background signal similar to the stray light background, it will need discrimination. In cases like that, it can help to place a bandfilter, which lets one line pass, plus a few nanometers of the background. If the background is the same with and without a filter, the spectrometer system provides good fidelity for the application. If a double monochromator is used that way, care must be taken because the double spectrometer might provide a better background suppression than the filter. If that happens, the bandfilter function would be the result and not the spectrometer. Of course, in many cases a second (possibly better) monochromator can be placed after the monochromator under inspection to analyze the output a second time for comparison. The efforts and costs to acquire the kind of equipment required for complex investigations can only be afforded by larger labs or by reference facilities. In any case, it is worth some expense to characterize the spectroscopy systems in use and to analyze the most critical part: the lower end of the wavelength range.

Figure 8.3 shows a system for the analysis of stray light in any optospectral system in the range of <200 to the NIR; it is also useful for spectral measurements of the relative transmission (see Fig. 7.7). The analysis system (right) consists of a pre-stage prism with a 25-cm focal length and a 50-cm multiple-grating mainstage. The internal stray light level in a distance of five bandwidths from the excitation band is flat and clearly below 10^{-6} . The detectors would be a PMT and a Si-InGaAs sandwich. The upper wavelength limit will depend on the InGaAs version: 1650, 2100, or 2600 nm. The system can be calibrated for radiometry, and besides of the sole application, it can be

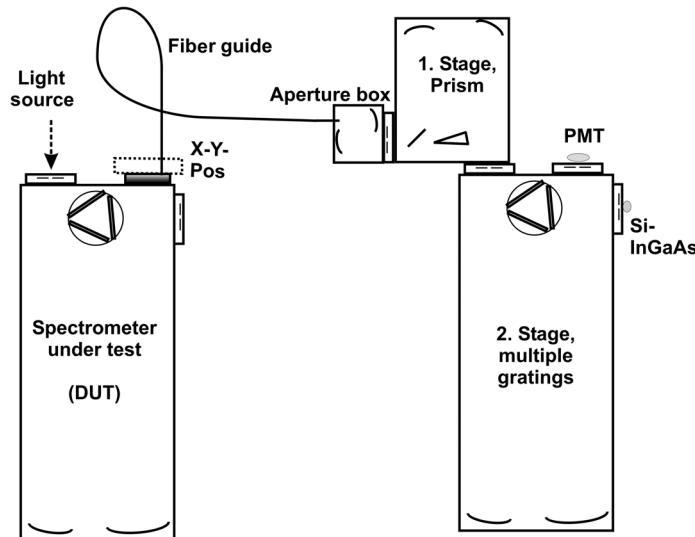


Figure 8.3 An optimized test setup to measure the stray light of spectrometers.

used as a spectrophotometer and for *limited* luminescence and Raman. It is coupled to the light source or the spectrometer under test (DUT) by optical fiber guides. These can be exchanged quickly and repeatedly for different applications. The wavelength coverage of the fibers will be 190–1100 nm or 400–2600 nm. The fiber cables are connected to an illumination box in front of the prism monochromator. The other end will be placed close to the exit slit of the DUT. Alternatively, it can be mounted to an x-y stage, which allows precise positioning in the focal plane of the output field of all kinds of building-block spectrometers. To avoid external light, a dark box will isolate the place of measurement from room light; to be adaptable, it uses flexible bellows. The dark box is not shown in the graph.

8.4 Minimizing the Impact of Disturbance through Optimization

Light that does not enter will not disturb the system. Unfortunately, that primitive rule is not easy to apply in broadband applications. Consider another example: performing excitation spectroscopy, such as luminescence, needs high intensity and high light density at the monochromator's exit. The preferred light source to achieve that is a xenon lamp, which is used often, especially when UV light is a matter of excitation. As discussed earlier (Section 6.4, Fig. 6.8), a typical xenon source emits approximately a factor of 10 of the median optical power above 400 nm compared to 200–250 nm. Because order sorting towards lower wavelengths is not required, there will normally be no filtering for excitation < 300 nm. The full xenon power enters. With regard to the stray light level in the excitation spectrometer's output, that is a difficult situation. If a monochromator of 1/3 m and a grating of

1200 mm^{-1} are used, stray light of up to 3% must be considered below 250 nm. If the application does not ultimately call for high power and density, the xenon source will be omitted, which also removes the troubles with the xenon peaks. The combination of separated D₂ and halogen lamps offers several options to limit the range of wavelengths present in the spectrometer. In the spectral range below 300 nm, where the disturbance is highest, only UV light enters the system. The stray light in the UV will drop to <1/10 compared to that produced by xenon, the difference caused only by the separation of the lamps. Even scattering plays no large role above 1 μm ; the energies from there can create disturbance, regardless. Reflections and thermal impact can create unwanted effects. As mentioned earlier, light outside the system will cause no problems. A short-pass filter, probably with a 1200-nm edge, will keep many energies out. Distilled water has that filter function, and it also removes heat. When using D₂ and halogen lamps, it is better to keep them separated than combine them in a single beam, even though it seems more comfortable. Only in applications that require parallel detection over the full wavelength range is a single beam advised.

The absolute share of stray light (from other wavelengths) at the actually measured wavelength is not the only point of decision. At least as important is the response of the receiver (sample, detector) upon the measured and the disturbing wavelengths. Consider one last example: An absorption experiment in the range of 200–1000 nm is acquired by a UV-Si array detector at once. The source is a combined D₂-halogen lamp. Based on the required dispersion, a spectrograph with 1% stray light at 200 nm, and 0.1% at 350 nm, is assumed. At 900 nm, the light source produces 20-fold power compared to 350 nm, and 10-fold compared to 200 nm. The detector efficiency is assumed to be 0.5 at 900 nm, 0.2 at 350 nm, and 0.05 at 200 nm. It is further assumed that all components, including the grating, treat all wavelengths equally well. If at 900 nm the final output signal is 10, it will be 0.5 at 350 nm, and 1 at 200 nm. The share of stray light at 350 nm will be 5%, at 200 nm it will be 2%, and above 400 nm it will quickly drop. For the absorbance measurements, these numbers mean that at 1 AU absorbance by the sample, the error at 350 nm is 50%, and at 200 nm it is 20%. If the system transfer curve is flattened by optimization of the grating, filtering, coating, and probably other parameters, in a way, the system considers 350 nm to be the best and 900 nm a little worse—the variation over the spectrum can be reduced from a factor of 20 to 3–5, which will linearly improve performance. If the whole range need not be acquired at once, the relative disturbance can be suppressed to <1%, allowing linear measurements up to 1.5 AU.

8.5 Reducing Stray Light

Improved surface quality of components reduces scattering. The frames, flanges, and edges of components should not be polished, blank, or sharp-edged to avoid the reflection of arriving (stray) light. The inner volume of a

spectrometer should be large because stray light reduces in an exponential manner with volume. UV light creates scattered signals, which can reach the detector even if no UV is required for the analysis. Because scattering grows with eV^4 , a signal at 250 nm creates ~ 250 times the scatter compared to a disturbance originating from 1 μm . Infrared light, in turn, is heat, which influences the thermal stability. The further the entering spectral interval reaches into the IR, the more heat it introduces. All kinds of distortions count if they reach the output. The interior of a spectrometer should be as empty as possible. The surface of walls and covers should be rough to reduce direct reflection. If constantly active electronic elements that create heat or light are found in the optical room, the apparatus might not be the best available choice. Modern systems incorporate stepper motors that can be mounted on the outside. The supporting electronics should sit in an isolated part of the housing.

If all parameters are carefully obeyed but stray light is still a problem, some provisions might be of assistance before a different spectrometer system must be obtained. The application of masks and baffles might help channel the beams and reduce scattered light. If possible, only the required wavelength interval should enter the spectrometer. Black razor-blade shields make for efficient baffles that can be placed on demand or where they have been tested best in a dark room. When affixing them, they should always be tilted versus the beam travel to reflect disturbing rays out. The standard black color might act reflective in the IR or be fluorescent in the UV. Special paints are available according to requirements. Black velvet might have the same drawbacks as paint. In the IR, they become slightly transmitting. Thin substrates, such as foils, with microscopic carbon tubes attached would be better outside the visible range. If a ray-tracing program is available, it could help identify which pathways the unwanted rays might travel, making it possible to stop them.

References

1. E. C. Fest, *Stray Light Analysis and Control*, SPIE Press, Bellingham, WA (2013) [doi: 10.1117/3.1000980].
2. W. Neumann, *Applications of Dispersive Optical Spectroscopy Systems*, SPIE Press, Bellingham, WA (to be published in 2014).

Chapter 9

Related Techniques

This chapter deals with spectrometer and component designs that do not fit the scope of this book but deserve a brief overview.

9.1 Compact, Fiber-Optic-Coupled Spectrographs

At the end of the 1980s, very compact spectrographs came to the market. Almost all of them were designed to be illuminated by a single optical fiber. The detector elements in the beginning were diode arrays 6.35 mm (256 pixels) or 12.7 mm (512 pixels) long. During the 1990s the upper limit in detector size grew up to 25.4 mm (1024 pixels). At the end of the century, CCD technology was added, mainly line-shaped. For approximate estimations, the known spectroscopic equations will produce satisfactory results, but keep in mind that the grating functions will be influenced by the curvature of the substrate. Thus, for accurate results, ray tracing is required. The longer the focal length of the system is, the better the approximation will fit. A large number of different suppliers, offering many different designs, can be found. The focal length ranges from 30–150 mm.

Stray light is a severe problem in compact systems. Even if everything inside the housing is fixed and optimized with light traps and baffles, a short optical pathway will allow unwanted rays to arrive at the detector. Because one of the important parameters for high dispersion is the focal length, both the dispersion and resolution are limited. If a compact spectrometer or a system based on it are considered, only one application should be applied. If no data exists, test measurements will determine if the performance is sufficient. Care should be taken when relying on the specifications provided by marketing materials. The bandwidth and resolution parameters are often mixed up, and the numbers might be provided for an unusually thin fiber cross-section. The linear range might only reflect a plane calculation of the ADC range, not taking background signals into account, and the stray light definition might be found true under conditions far different than the planned

measurements. Research and standardization, along with any kind of variation, are definitely out of reach for that family of optical systems.

These warnings do not mean that the data are not useful. They are indeed if the fidelity of the data produced matches the result, but the spectrum provided by an internal standard (whatever the measurement) must meet pre-defined values within a certain tolerance. That means, in turn, that there are applications in production processes, quality control, and real-time spectral tracing ready to be solved by compact systems after serious evaluation has occurred.

The most efficient version uses only one reflecting surface, which combines the refocusing mirror with a holographic grating. Because some standard fiber materials provide a full cone angle of 25 deg ($f/2.3$, $\alpha = 0.216$), the optical component is designed to fit at a distance of typically 30–150 mm. The distance to the detector (focal length) is the same, in most cases. Figure 9.1 shows both distances to be 50 mm, with a detector length of 25.4 mm. On the right, a system with two optical components is shown. It has the advantage that the grating does not need to capture the light, disperse it, and also refocus it. That gives both the mirror and the grating a “normal” concave shape. This increases the possibility that both elements could be obtained from different manufacturers, whereas in the single-optics case, custom products are required. The fiber cable is typically equipped with a single fiber whose cross-section is

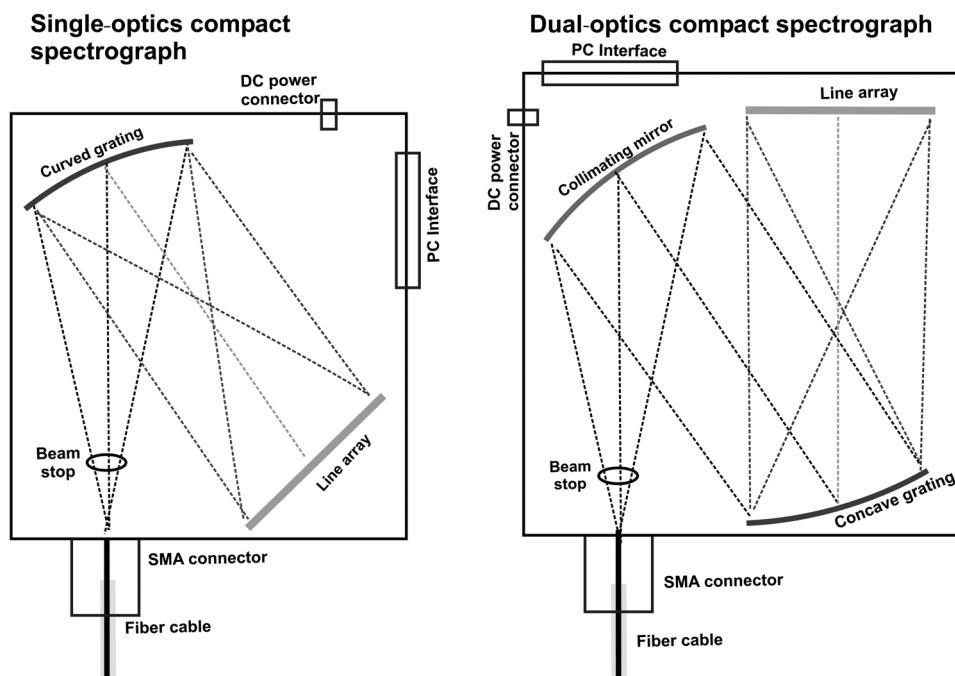


Figure 9.1 Two popular designs of compact spectrometers.

critical. The connector is most often an SMA. In the case shown here, the end of the SMA is the outside of the wall of the housing, which also represents the focal plane of the mirror and is compatible with the SMA design. Many instruments feature a circular or squared beam stop to avoid over-illumination if the fiber's emission angle exceeds that of the instrument. Some instruments feature a lens between the fiber and first reflection, which also limits the cone angle and helps keep the optical elements smaller at the cost of optical quality.

If the focal length is 50 mm, the optical parts need to provide at least a 20-mm diameter, or $\Omega = 0.125$. If the active detector length is 25.4 mm, the following approximate dispersion and coverage values can be expected:

- 300-mm⁻¹ grating: reciprocal dispersion (rd) = 50 nm/mm, coverage (c) = 1300 nm, and upper limit (ul) = 5000 nm;
- 1200-mm⁻¹ grating: rd = 12 nm/mm, c = 300 nm, ul = 1300 nm;
- 1800-mm⁻¹ grating: rd = 7 nm/mm, c = 175 nm, ul = 850 nm;
- 2400-mm⁻¹ grating: rd = 2.2 nm/mm, c = 55 nm, ul = 650 nm;
- 3600-mm⁻¹ grating: rd = 1.5 nm/mm, c = 37 nm, ul = 420 nm.

Because the gratings are holographic in any case, an almost infinite number of line densities and optimization wavelengths can be produced. The spectral bandwidth, guided to a single pixel or vertical channel, depends on the pixel size. The pixel width can vary between 5–50 μm , depending on the detector model. Calculation is easy because it is the coverage divided by the number of channels. Defining the resolution is difficult and almost cannot be calculated because of the influencing parameters; theoretically it would be 3 times the bandwidth per channel, but that would imply that no aberrations exist, which is not expected in short systems. A factor of 5 is probably more realistic but only if the fiber cross-section is between a factor of 1.3 and 2 of the channel width. Assuming a channel width of 25 μm , a 50- μm fiber would be perfect. Now, with a grating of 1200 mm⁻¹, the resolution can be expected to be 1.2 nm. In favor of higher light throughput, the fiber cross-section is often larger, probably 200 μm . That setup will effect a resolution of probably 5 nm, with the advantage that aberrations will play no or only a minor role. Whether the order overlay is dangerous as soon as more than one octave is recovered depends on the application and the required fidelity. Filtering is probably required at the front end of the fiber.

To correct that problem, some suppliers offer an order-reducing filter coating for the array. There are spectrometers with and without detector TE cooling and built-in ADC. If no cooling is integrated but the ADC is, USB communication is state-of-the-art and makes an extra power supply obsolete. Additional cooling will require extra supply. Other systems feature a parallel link to the PC that allows faster read-out and control, but it is more complicated. The versions in between are manifold. A compact spectrometer with a 50-mm focal length can be as small as 60 mm \times 60 mm \times 30 mm;

the light source would not be much larger, making the whole system very attractive. Si arrays typically cover 350–1050 nm, or 200–1050, while InGaAs detectors are used for 750–1650 nm or 750–2100 nm, or even 750–2600 nm.

9.2 Programmable Gratings

Based on emerging MEMS (microelectromechanical systems) technology, the first programmable gratings have been presented on a R&D scale. Silicon growing and etching allows for the production of parallel structures several millimeters in length, with a distance of a few microns between the lines. A light-transmitting, electricity-conducting layer is placed above an elastic reflecting layer, which itself is placed on a substrate. The upper layer is equipped with pillars, which in turn are riding on the second layer. If a static voltage is applied to the upper layer, it moves towards the substrate. That arrangement creates a sine-wave-shaped modulation in the second layer that is similar to the shape of the lines of a holographic grating. If only a small voltage is applied, the modulation is low, and the optimal dispersion is towards shorter wavelengths. With increasing voltage, the modulation amplitude increases, and the dispersion efficiency is better at longer wavelengths. Only samples for the NIR–IR range have been produced, depending on the line frequencies and modulation depth achieved; however, it is an emerging technology worth watching.

9.3 Bragg Gratings and Filters

Bragg's rules allow one to "write" a holographic layer structure of alternating refractive indices between the core and cladding (see Section 6.7) into a cylinder of optical fiber material. The 3D structures can be programmed such that a precisely defined wavelength will experience a 180-deg phase change at every reflection, which means that it is dimmed each time. After many reflections, the transmission of that particular wavelength tends towards zero, producing a very efficient notch filter (Fig. 9.2). In the literature, devices like this are called "Bragg gratings" even though they work more like very narrow negative filters. Because the structures require that the light travels under exact angles, the Bragg area follows a fiber bolt of the same material combination to adjust the angles before the light reaches the filter zone. Bragg devices in the shape of standard transmission filters are similar. 3D holographic structures are burned by a laser application into a polymer substrate. At a precisely defined wavelength, the light is totally reflected, while all other wavelengths pass with little or no suppression. For correct operation, the filter requires perfectly collimated light; to that end, notch filter systems include a lens system or camera objective before and after the filter. Polarized light of a defined angle is preferred for the best efficiency. The notch wavelength can be tuned within a narrow wavelength range by turning the filter. In applications that require the strong suppression of a single

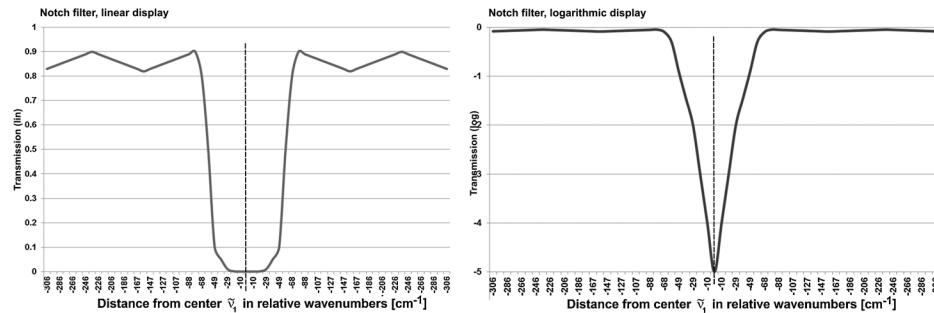


Figure 9.2 Bragg filter function (notch filter).

wavelength (Raman or luminescence), the Bragg filter is very efficient and helpful. Manufacturers also call Bragg filters "holographic notch filters." The best products provide bandwidths of 100 cm^{-1} or even less at FWHM.

Bragg gratings and filters are extremely useful for suppressing a single wavelength within a spectrum. The damping reaches up to 10^5 or even 10^6 . At 500 nm , 100 cm^{-1} is equivalent to 2.5 nm ; because the technology can only be applied to certain materials, the range between ~ 350 – 1300 nm is covered.

A typical notch curve is shown in Fig. 9.2. In May 2012, a US patent was filed that describes the enhanced capabilities of the Bragg filter technique. A company has succeeded in stacking several plane Bragg filters, producing a "volume holographic grating" (VHG). According to the literature,¹ the resulting notch filters provide FWHM values as small as 0.4 nm , or 20 cm^{-1} , providing line suppression of up to 10^6 , or 6 AU.

9.4 Hadamard Spectrometer

The Hadamard transformation in optical spectroscopy, in simplified terms, is a competitive version of the Fourier transformation (FT). Whereas FT spectroscopy instruments depend on the phenomenon of interference, Hadamard spectrometers work with dispersed light. The Hadamard algorithm³ works with all amounts of data points so long as they are defined in advance. Because the result is drawn from a matrix that provides summation and subtraction of the recorded channels upon a shift of one channel in between, the number of passes (typically spectral scans) must be at least the same as the number of channels in the matrix.

A look at the past shows the power of the Hadamard transform before computer power was available. Until the 1970s, analog FT instruments had extremely complicated electronics. The gathered data had to be calculated by external computer systems to produce the final results, whereas Hadamard instruments provided the results in almost real time. Conversely, many applications that were historically realized by Hadamard systems are now solved either by the fast Fourier transform (FFT) or parallel detection, with the help of personal computers.

Still, Hadamard systems offer some advantages in certain situations. The methods of profitably applying Hadamard algorithms are wider spread than it might seem at first. The advantages include the following: common signals (such as background) are automatically removed; a single-channel detector can receive discrete signals without changing the disperser position during

Table 9.1 Hadamard matrix for the reduction of six data channels.

Channel No.	1	2	3	4	5	6
Scan (1)	+	+	+	+	+	+
Scan (2)	+	+	+	-	-	-
Scan (3)	+	+	-	-	+	-
Scan (4)	+	+	-	-	-	-
Scan (5)	+	-	-	-	+	+
Scan (6)	+	-	+	-	+	-

acquisition; a rather large output area can be used to obtain high luminosity; and the data are fully compatible with those of monochromator systems (FT data are different). In comparison to lock-in technology, the efficiency is higher because the light flux is continuously detected. During operation, a Hadamard spectrometer must modulate the spectral information, which can happen through a sector disk or a spiral disk (as demonstrated in Fig. 9.3). Some authors have described solutions with vibrating or scanning slit mechanisms, fast scanning, or rotating gratings, which produce the same result in principle.

9.4.1 Principle of Hadamard measurements

During acquisition, the data stream is uninterrupted. Steering electronics store the different data in a pre-defined number of memories. The Hadamard transformation requires at least as many repetitions as the data points that are defined. Routine Hadamard applications include the recording of polarization behavior, as in ORD, CD, or ellipsometry.² They include either rotating polarizers or sine modulation through all states of polarization. The measurement requires four data points per analysis, which are contiguously spread over one polarization cycle. Thus, the transformation requires at least four cycles. A spectral measurement (of irrelevant polarization) with six spectral points requires the Hadamard matrix shown in Table 9.1 for reduction. “+” means that the data are additive, and “-” means that the sign is inverted before addition. This leads to positive and negative data, which automatically subtracts constant signals (background) out of the result.

9.4.2 Hadamard setups

Hadamard arrangements can look very different because of the wide range of possible applications. Figure 9.3 shows two versions of a spectrograph. The spectrograph itself works like a standard one for parallel detection. The Hadamard encoder is placed in the focal plane or very close to it. In the example, the field is 40 mm wide and 20 mm high. In the case shown, the encoder is a rotating disk. The passing light signals will be captured by a single detector at least as large as the field (PMT and silicon detectors of that size are standard). In the figure, the Hadamard encoder is drawn in two versions; on the left is a segmented disk for six wavelengths. While the disk rotates

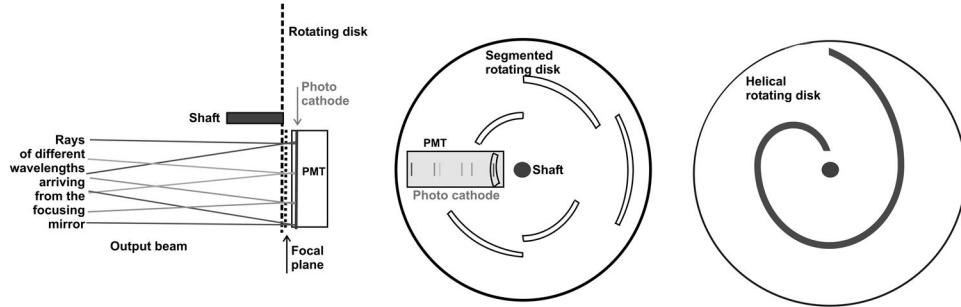


Figure 9.3 Output section of a Hadamard spectrograph.

constantly, a trigger system designates the different memories of data storage. The triggers can be optical, actuated through little holes in the outer frame of the disk (not shown in the graph). One wavelength at a time is detected, but no interruption happens in photon flux. After six turns, the set of data would be complete, but more turns improve the SNR. The illumination time is the same for all wavelengths because the angular speed is the same. The sectors can be of different widths to vary the bandwidth. Different sets of wavelengths will require different disks. The image on right in Fig. 9.3 is a spiral-shaped disk that recovers the whole spectral interval. This requires more memories, or else the bandwidth per channel would be rather wide. Because the algorithm calls for a contiguous data flow, the required resolution defines the width of the spiral and the number of data points. In any case, the “slit” formed by the spiral will wipe over the center wavelength during exposure of a data channel. That behavior leads to averaging, which might not be a disadvantage. In both examples, the speed of rotation can be varied.

References

1. C. Moser and G. Steckman, "Filters to Bragg About: Volume holographic gratings offer distinctive filter qualities," *Photonics Spectra* (June 2005).
2. S. Agaian, H. Sarukhanyan, K. Egiazarian, and J. Astola, *Hadamard Transforms*, SPIE Press, Bellingham, WA (2011) [doi:10.1117/3.890094].

Index

A

aberration, 24, 33, 48, 51–52, 54, 55, 57, 59–61, 92, 127, 129–131, 135
absorption, 26, 96, 147
acceptance angle, 144
active surface, 29, 104, 201
additive dispersion, 115
additive setup, 39, 108, 112
adjustment, 20, 22, 107, 118
aluminum, 66, 80, 127, 136
aluminum oxide, 80
angle
 at the mirror, 15
 dispersion, 17, 19, 44, 48
 grating, 21, 37, 57
 illumination, 22, 74, 96
 internal, 14, 23–25, 27, 30, 52, 96
aperture, 13, 29, 32, 53, 105, 111
astigmatism, 23, 55
asynchronous noise, 167

B

ball-shaped radiation, 156, 208
bandwidth
 electrical, 8, 10, 171
 optical, 4, 33, 39
border frequency, 7

C

calcium fluoride (CaF_2), 130
calibration, 102
cathode, 147, 149–151, 155
charge, 153, 157, 159, 171, 175, 200

coating, 79, 80

collimating optics, 13
coma, 54, 57–58, 131
cone-shaped radiation, 13, 94, 111, 207
contrast, 58, 60, 66, 107, 120
conversion, 3–4, 92, 152, 176, 195
cooling, 137, 144, 149, 151, 159, 161, 172, 175
cosecant, 101
cosine, 37, 100
counting, 151, 157
cross-section, 189, 210, 236, 271
current, 7, 147–148, 150, 188
curved slit, 20, 22, 28, 52, 57
Czerny–Turner, 23, 25, 87–88, 91, 119, 133–134

D

damping, 6–7, 63, 168
deep depletion, 184, 186
deep UV, 26, 98, 198
deuterium (D_2), 214, 217–218
dispersion, 16–17, 36–37, 39, 74, 89–90, 96, 100, 120, 245, 252
distortion, 24, 80
distribution, 10, 37, 43, 48, 86, 110, 173, 207, 252
drift, 61, 63, 120, 127, 136

E

Ebert–Fastie, 20, 22, 25, 55, 87, 119, 133
electron volt (eV), 2–3

entrance slit, 33, 37, 65, 254

exit aperture, 13

exit slit, 37, 42, 45, 262

exponential functions, 6

F

f-number, 14, 29, 31–32, 36, 241

fluorescence, 108, 137, 235

focal length, 5, 29, 44, 48, 51, 58, 109, 239, 242, 269

focusing mirror, 30, 38, 87, 92

free spectral range (FSR), 16, 65

frequency, 2–4, 7–8, 161, 163, 171, 175

front illumination, 183, 186, 201

G

gain, 143, 153, 170

general rule, 5

ghost, 44, 58, 79

glass, 61, 128, 136, 151, 235

Globar, 218, 225

gold, 80, 212, 220, 226

grating, 1, 14–15, 21, 22

curved, 25, 27, 71, 73, 94

Echelle, 71–72, 78, 99, 122–123, 128, 132

holographic, 69, 71, 77, 79, 98

rotation, 84–85, 100, 115

ruled, 71, 77, 79, 122

grism, 75–76

H

hollow cathode lamp (HCL), 207

I

image correction, 52, 56

image magnification, 214

image quality, 20, 51

indium-gallium-arsenide (InGaAs), 158, 171, 185

integration, 8, 45, 253

Invar, 61, 102, 136–137

L

light-emitting diode (LED), 207, 213, 253, 259

liquid nitrogen (LN), 159, 161, 172, 198

Littrow, 17–18, 21, 25, 55, 60, 70, 132

lock-in, 163–169

loss, 235, 237

luminescence, 233, 266

luminosity, 29, 31–32, 67, 119

M

magnesium fluoride (MgF_2), 130, 226

modulation, 60, 124, 155, 158,

162–163, 272

monochromator, 4

N

Nernst, 225

number, 14, 42

O

off-axis, 54, 103, 114, 249

optical filter, 9, 16, 64, 271

optimization, 53, 136, 193, 228, 256, 260, 266

P

photon energy, 96, 100, 148, 245

pinhole, 13, 17, 132

Planck's law, 159, 217

polarization, 2, 69–70, 76, 78, 235, 274

polychromator, 4

prism, 13, 16, 17, 22, 64–65, 73–76, 95–96, 99

pulse, 150, 154, 157, 198, 203

Q

quartz, 26, 130, 232

R

radiometry, 231, 253, 263

radius, 14, 28, 52, 234

Raman, 2, 101, 108, 118, 137, 235

read-out, 135, 171, 175, 177–183, 187, 189, 191–193
rear-side illumination, 184, 186
reflection, 59, 80, 213, 222, 256
refractive index (RI), 97, 231, 272
resolution, 39–40, 42, 44–45, 47–49, 53, 105, 119

S

sagittal focus, 23, 54–55
sapphire, 130, 226
scatter, 59, 79, 260, 268
sensor, 171, 174, 189, 192, 200
separation, 113, 130–131, 158, 187
signal-to-noise ratio (SNR), 146–148, 159, 161, 167, 228
silicon, 158, 169, 171, 183, 198
sine, 90, 100, 165, 246, 247, 250
sine wave, 7, 71, 165
spectral line, 5, 46, 66, 249, 261, 265
spectral order, 15, 63
spectrograph, 4, 26, 35, 37, 45, 47, 49, 53, 55, 60, 73, 83, 123, 269, 274
spectrometer, 1, 4, 13–14
stability, 60, 62, 120, 136
standard deviation (STD), 144, 146, 254

steady state, 162
steradian (sr), 162, 209
stray light, 58–60, 259–266, 268, 269
subtractive, 39, 44, 108, 113–114, 117, 119–121
symmetry, 21, 22, 24, 28, 56, 85, 87
synchronization, 162, 169, 187

T

tangential focus, 23, 54–55, 59–60
taper, 155, 173, 194, 238
transmission, 29, 33, 36, 73, 75, 99, 110, 225
tungsten halogen, 77, 217, 219

V

vacuum, 3, 25, 216
vignetting, 38
voltage, 7, 9, 144, 147, 154, 158, 178, 272

W

wavelength, 2–4
wavenumber, 2, 4, 95

X

xenon, 216, 218, 227, 266



Wilfried Neumann was born in 1945 in a small township south of Frankfurt, Germany. After graduating from high school, he was educated in radio and TV electronics. After working in the R&D department of Hottinger-Baldwin-Messtechnik GmbH (in the field of mechanical stress analysis), he attended a college for electronics, high-frequency technology, and computer hardware and software. He earned a degree as an approved technician in 1973.

Neumann based his career on signal recovery and optical spectroscopy. He worked in the service, application, sales, and marketing of several pertinent international companies. Early on, he worked in service and applications at Perkin-Elmer (analytical instruments) and Varian (optical spectroscopy, now part of Agilent). Between 1977 and 1989, he was an area sales representative at EG&G Instruments (signal recovery and optical spectroscopy). In 1989, he managed the new division of optical spectroscopy at LOT-Oriel GmbH, a distribution company in the field of technical optics and lasers. One of the companies he marketed in Germany was SOPRA, Inc. (Paris, France), a spectroscopy company; the division developed quickly, and in 1991, Neumann became the founder, shareholder, and CEO of a new branch called SOPRA GmbH. In 2000, he joined Roper Scientific GmbH, the German branch of Roper's CCD and optical imaging and spectroscopy business, headed by Princeton Instruments. He served as area sales and product manager for special, research-oriented spectroscopy systems.

After his retirement from regular work in 2005, Neumann started a consulting office, offering spectrometer system maintenance and service, as well as consultation, justification, design, and construction for optical spectroscopy and radiometry. To support this endeavour, and because comprehensive literature on the basics of dispersive optical spectroscopy did not exist, he created a website (www.spectra-magic.de) that has since become a popular online resource.

Fundamentals of Dispersive Optical Spectroscopy Systems

Wilfried Neumann

Bridging the gap between basic theoretical texts and specific system recommendations, *Fundamentals of Dispersive Optical Spectroscopy Systems* addresses the definition, design, justification, and verification of instrumentation for optical spectroscopy, with an emphasis on the application and realization of the technology. The optical spectroscopy solutions discussed within use dispersive spectrometers that primarily involve diffraction gratings. Topics include dispersion elements, detectors, illumination, calibration, and stray light. This book is suitable for students and for professionals looking for a comprehensive text that compares theoretical designs and physical reality during installation.

ISBN 9780819498243



SPIE.

P.O. Box 10
Bellingham, WA 98227-0010

ISBN: 9780819498243
SPIE Vol. No.: PM242