Fourier-Transform Spectroscopy Instrumentation Engineering

Tutorial Texts Series

- *Fourier-Transform Spectroscopy Instrumentation Engineering*, Vidi Saptari, Vol. TT61
- *The Power- and Energy-Handling Capability of Optical Materials, Components, and Systems*, Roger M. Wood, Vol. TT60
- *Hands-on Morphological Image Processing*, Edward R. Dougherty, Roberto A. Lotufo, Vol. TT59
- *Integrated Optomechanical Analysis*, Keith B. Doyle, Victor L. Genberg, Gregory J. Michels, Vol. TT58
- *Thin-Film Design: Modulated Thickness and Other Stopband Design Methods*, Bruce Perilloux, Vol. TT57
- *Optische Grundlagen für Infrarotsysteme*, Max J. Riedl, Vol. TT56
- *An Engineering Introduction to Biotechnology*, J. Patrick Fitch, Vol. TT55
- *Image Performance in CRT Displays*, Kenneth Compton, Vol. TT54
- *Introduction to Laser Diode-Pumped Solid State Lasers*, Richard Scheps, Vol. TT53
- *Modulation Transfer Function in Optical and Electro-Optical Systems*, Glenn D. Boreman, Vol. TT52
- *Uncooled Thermal Imaging Arrays, Systems, and Applications*, Paul W. Kruse, Vol. TT51
- *Fundamentals of Antennas*, Christos G. Christodoulou and Parveen Wahid, Vol. TT50
- *Basics of Spectroscopy*, David W. Ball, Vol. TT49
- *Optical Design Fundamentals for Infrared Systems, Second Edition*, Max J. Riedl, Vol. TT48
- *Resolution Enhancement Techniques in Optical Lithography*, Alfred Kwok-Kit Wong, Vol. TT47
- *Copper Interconnect Technology*, Christoph Steinbrüchel and Barry L. Chin, Vol. TT46
- *Optical Design for Visual Systems*, Bruce H. Walker, Vol. TT45
- *Fundamentals of Contamination Control*, Alan C. Tribble, Vol. TT44
- *Evolutionary Computation: Principles and Practice for Signal Processing*, David Fogel, Vol. TT43
- *Infrared Optics and Zoom Lenses*, Allen Mann, Vol. TT42
- *Introduction to Adaptive Optics*, Robert K. Tyson, Vol. TT41
- *Fractal and Wavelet Image Compression Techniques*, Stephen Welstead, Vol. TT40
- *Analysis of Sampled Imaging Systems*, Richard H. Vollmerhausen, Ronald G. Driggers, Vol. TT39
- *Fundamentos de Electro-Optica para Ingenieros*, Glenn D. Boreman, Vol. TT37
- *Infrared Design Examples*, William L. Wolfe, Vol. TT36
- *Fundamentals of Machine Vision*, Harley R. Myler, Vol. TT33
- *Design and Mounting of Prisms and Small Mirrors in Optical Instruments*, Paul R. Yoder, Jr., Vol. TT32
- *Basic Electro-Optics for Electrical Engineers*, Glenn D. Boreman, Vol. TT31
- *Optical Engineering Fundamentals*, Bruce H. Walker, Vol. TT30
- *Introduction to Radiometry*, William L. Wolfe, Vol. TT29
- *Lithography Process Control*, Harry J. Levinson, Vol. TT28
- *Introduction to Interpretation of Graphic Images*, Sergey V. Ablameyko, Vol. TT27
- *Introduction to Imaging Spectrometers*, William L. Wolfe, Vol. TT25
- *Introduction to Infrared System Design*, William L. Wolfe, Vol. TT24
- *Introduction to Computer-based Imaging Systems*, Divyendu Sinha, Edward R. Dougherty, Vol. TT23
- *Optical Communication Receiver Design*, Stephen B. Alexander, Vol. TT22
- *Mounting Lenses in Optical Instruments*, Paul R. Yoder, Jr., Vol. TT21
- *Introduction to Real-Time Imaging*, Edward R. Dougherty, Phillip A. Laplante, Vol. TT19
- *Integration of Lasers and Fiber Optics into Robotic Systems*, Janusz A. Marszalec, Elzbieta A. Marszalec, Vol. TT17
- *An Introduction to Nonlinear Image Processing*, Edward R. Dougherty, Jaakko T. Astola,Vol. TT16
- *Introduction to Optical Testing*, Joseph M. Geary, Vol. TT15
- *Image Formation in Low-Voltage Scanning Electron Microscopy*, Ludwig Reimer, Vol. TT12
- *Diazonaphthoquinone-based Resists*, Ralph R. Dammel, Vol. TT11
- *An Introduction to Optics in Computers*, Henri H. Arsenault, Yunlong Sheng, Vol. TT08
- *Digital Image Compression Techniques*, Majid Rabbani, Paul W. Jones, Vol. TT07

Fourier-Transform Spectroscopy Instrumentation Engineering

Vidi Saptari

Tutorial Texts in Optical Engineering
Volume TT61

Arthur R. Weeks, Jr., Series Editor

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

Library of Congress Cataloging-in-Publication Data

Saptari, Vidi.
Fourier-transform spectroscopy instrumentation engineering / Vidi Saptari.
p. cm. — (SPIE tutorial texts ; v. TT61)
Includes bibliographical references and index.
ISBN 0-8194-5164-9 (soft cover)
1. Fourier-transform spectroscopy–Instruments. I. Title. II. Series.

QC454.F7S37 2003
621.36'1–dc22 2003055705
CIP

Published by

SPIE—The International Society for Optical Engineering
P.O. Box 10
Bellingham, Washington 98227-0010 USA
Phone: 360.676.3290
Fax: 360.647.1445
Email: spie@spie.org
www.spie.org

Printed in the United States of America.

Introduction to the Series

Since its conception in 1989, the Tutorial Texts series has grown to more than 60 titles covering many diverse fields of science and engineering. When the series was started, the goal of the series was to provide a way to make the material presented in SPIE short courses available to those who could not attend, and to provide a reference text for those who could. Many of the texts in this series are generated from notes that were presented during these short courses. But as stand-alone documents, short course notes do not generally serve the student or reader well. Short course notes typically are developed on the assumption that supporting material will be presented verbally to complement the notes, which are generally written in summary form to highlight key technical topics and therefore are not intended as stand-alone documents. Additionally, the figures, tables, and other graphically formatted information accompanying the notes require the further explanation given during the instructor's lecture. Thus, by adding the appropriate detail presented during the lecture, the course material can be read and used independently in a tutorial fashion.

What separates the books in this series from other technical monographs and textbooks is the way in which the material is presented. To keep in line with the tutorial nature of the series, many of the topics presented in these texts are followed by detailed examples that further explain the concepts presented. Many pictures and illustrations are included with each text and, where appropriate, tabular reference data are also included.

The topics within the series have grown from the initial areas of geometrical optics, optical detectors, and image processing to include the emerging fields of nanotechnology, biomedical optics, and micromachining. When a proposal for a text is received, each proposal is evaluated to determine the relevance of the proposed topic. This initial reviewing process has been very helpful to authors in identifying, early in the writing process, the need for additional material or other changes in approach that would serve to strengthen the text. Once a manuscript is completed, it is peer reviewed to ensure that chapters communicate accurately the essential ingredients of the processes and technologies under discussion.

It is my goal to maintain the style and quality of books in the series, and to further expand the topic areas to include new emerging fields as they become of interest to our reading audience.

Arthur R. Weeks, Jr.
University of Central Florida

Contents

Preface / xi

Chapter 1
Spectroscopy Instrumentation / 1
1.1 Introduction / 1
1.2 Types of Spectrometers / 1
1.2.1 Dispersive spectrometers / 1
1.2.2 Filter-based spectrometers / 2
1.2.3 Fourier-transform spectrometers / 3
1.3 Advantages of FT Spectrometers / 4
1.3.1 Throughput or Jacquinot advantage / 4
1.3.2 Multiplex or Felgett advantage / 6
1.4 Discussions on FT Spectrometer Advantages / 8
References / 9

Chapter 2
Signal-to-Noise Ratio / 11
2.1 Signal-to-Noise Ratio Defined / 11
2.2 Quantifying Signal-to-Noise Ratio / 12
2.3 Practical Considerations / 15

Chapter 3
Principles of Interferometer Operation / 17
3.1 Overview / 17
3.2 Quantitative Explanation / 19
3.2.1 Light as a wave / 19
3.2.2 Measurable light quantity / 21
3.2.3 Interference and superposition / 22
3.2.4 Polychromatic source / 24
3.2.5 Fourier-transform routine / 25
3.3 Theoretical Resolution / 26
3.3.1 Retardation distance / 26
3.3.2 Divergence angle / 29
3.4 Interferogram Digital Processing / 32
3.4.1 Apodization / 32

3.4.2 Zero filling / 34
3.4.3 Phase correction / 34
References / 34

Chapter 4
Interferometer Alignment Errors / 37
4.1 Error Characteristics / 39
4.2 Interferogram-Modulation Error / 40
4.3 Interferogram-Sampling Error / 45
References / 46

Chapter 5
Motion Components and Systems / 47
5.1 Actuators / 48
5.1.1 Actuator types / 48
5.1.2 Performance specifications for actuators / 48
5.2 Driver and Amplifier / 50
5.3 Bearings and Coupling Mechanism / 50
5.3.1 Bearing types / 51
5.3.2 Performance specifications for bearings / 51
5.3.3 Discussion on bearings / 53
5.4 Position Sensor / 55
5.4.1 Laser interferometer / 56
5.4.2 Sensor specifications / 58
5.5 Homodyne Interferometer Sources of Error / 59
5.5.1 Source wavelength variations / 59
5.5.2 Refractive index variations / 59
5.5.3 Laser detector errors / 60
5.5.4 Polarization errors / 60
5.5.5 Alignment errors / 61
5.6 Actuator, Sensor, and Mirror Mounting / 61
5.6.1 Actuator location vs. center of mass / 61
5.6.2 Sensor location vs. mirror location / 62
5.6.3 Sensor location vs. actuator location / 62
5.7 Digital Motion Controller / 63
References / 64

Chapter 6
Interferogram Data Sampling / 65
6.1 Step Scan vs. Continuous Scan / 66
6.2 Sampling Period / 67
6.2.1 The absolute minimum: the Nyquist criteria / 67
6.2.2 More than the Nyquist minimum / 68
6.2.3 Anti-alias filter / 69
6.3 Accuracy with Respect to OPD / 71
6.3.1 Effects of sampling errors to the output spectrum / 71
6.3.2 Sources of sampling error / 73

6.4 Repeatable Clock Position / 74
6.5 Hardware Considerations / 74
References / 76

Chapter 7
Data Acquisition / 77
7.1 DAQ Hardware Formats / 77
7.2 Analog Inputs / 77
7.2.1 Number of channels / 78
7.2.2 Sampling rate / 78
7.2.3 Resolution / 78
7.2.4 Input ranges / 78
7.2.5 Accuracy / 79
7.3 Hardware Triggering and Clock Signal / 79
7.4 Effects of ADC Resolution on Spectral SNR / 80

Chapter 8
The Detector / 83
8.1 Noise-Equivalent-Power / 83
8.2 Spectral SNR quantification / 84
8.3 Detector Types / 87
8.4 Selection Guidelines / 88
8.4.1 Spectral range and responsivity / 88
8.4.2 Active area / 88
8.4.3 Noise and detectivity /89
8.4.4 Linearity / 89
8.4.5 Temporal bandwidth / 90
8.4.6 Summary / 90
References / 91

Chapter 9
Consideration of Optics and Interferometer Alignment / 93
9.1 System Throughput / 93
9.2 Lenses vs. Mirrors / 95
9.3 Interferometer Alignment Procedure / 96

Chapter 10
Signal-to-Noise Ratio Enhancement Techniques / 99
10.1 Identification of Error Sources / 99
10.1.1 Detector noise / 99
10.1.2 Digitization noise / 100
10.1.3 Interferometer alignment error / 101
10.1.4 Sampling error / 101
10.1.5 Light-source variation / 101
10.2 Averaging / 102
10.3 Temporal Modulation and Bandwidth Narrowing / 103

10.4 Spectral-Bandwidth Narrowing / 104
10.5 Spectral Post-Processing / 105
10.6 Double Beaming / 106
10.6.1 Electrical subtraction / 107
10.6.2 Optical subtraction / 107
10.7 Gain ranging / 109
10.8 Dynamic Tilt Compensation / 109
References / 110

Appendix A
Simulation of Static-Tilt Error / 111

Appendix B
Sampling Circuit Example / 113

Appendix C
Simulation of Sampling Error / 115

Index / 117

Preface

This book covers the engineering aspects of Michelson-interferometer-based Fourier-transform (FT) spectrometers. While other Fourier-transform spectroscopy textbooks have generally focused on the theoretical aspects and applications, this book provides insights to the design and instrumentation of FT spectrometers. This book is intended for scientists and engineers who wish to understand the operation of the instrument to the extent where they have the knowledge to improve results through hardware, software and/or procedural modifications. This book is also a proper introductory text for an engineer who plans to design a custom FT spectrometer.

Many applications today require the FT spectrometer to perform close to its limitations, such as taking many quantitative measurements in the visible and in the near-infrared wavelength regions. In such cases, the instrument should not be considered as a perfect "black box." Knowing where the limitations of performance arise and which components must be improved are crucial to obtaining repeatable and accurate results. One of the objectives of this book is to help the user identify the instrument's bottleneck.

The book begins with an overview of the types of spectrometers and the fundamental advantages of FT spectrometers (Chapter 1). Chapter 2 is devoted to clearly defining signal-to-noise ratio (SNR), a performance figure extensively used throughout this book. In Chapter 3, details of the interferometer operation are discussed. The design and limitations of the subsystems and components that include the motion components, the analog-to-digital (A/D) conversion process, the detection system, and the optical components are covered in Chapters 4 through 9. Chapter 10 introduces various practical techniques to identify the limiting source of error, and techniques to improve the obtainable SNR.

I am indebted to Professor Kamal Youcef-Toumi at Massachusetts Institute of Technology (MIT) for giving me the opportunities to be involved in various first-rate R&D projects during my time at MIT, one of which led me to the exciting world of FT spectroscopy. I deeply thank Merry Schnell of SPIE, the main editor of this project, for her thoroughness and professionalism. Bradley M. Stone and David Ball did an excellent job in reviewing the original manuscript. Rick Hermann and Timothy Lamkins of SPIE provided generous assistance from the early stage of the project. All have my deepest appreciation.

Vidi Saptari
September 2003

Chapter 1
Spectroscopy Instrumentation

1.1 Introduction

Fourier-transform spectrometers (FTS or FT spectrometers) have been replacing the dispersive instruments in many infrared and near-infrared applications over the last couple of decades. Their inherent advantages compared with the dispersive instruments are proven and well accepted by scientists and engineers working in the field of spectroscopy. This chapter provides a general overview of the types of spectrometers commonly used today, focusing on the differences in their operating principles. An analysis of an FT spectrometer's advantages over its dispersive counterpart is then presented.

1.2 Types of Spectrometers

Spectrometers can be categorized into three main types based on their principles of operation: dispersive, filter-based, and Fourier-transform instruments.

1.2.1 Dispersive spectrometers

As the name suggests, dispersive spectrometers generate spectra by optically *dispersing* the incoming radiation into its frequency or *spectral* components, as illustrated in Fig. 1.1. Common dispersive elements include prisms and gratings. Dispersive spectrometers can be further classified into two types: monochromators and spectrographs. A monochromator uses a single detector, narrow slit(s) (usually two, one at the entrance and another at the exit port), and a rotating dispersive element allowing the user to observe a selected range of wavelength. Figure 1.1 shows the simplified schematic of a monochromator.

A spectrograph, on the other hand, uses an array of detector elements and a stationary dispersive element. In this case, the slit shown in the figure is removed, and spectral elements over a wide range of wavelengths are obtained at the same time, therefore providing faster measurements with a more expensive detection system.

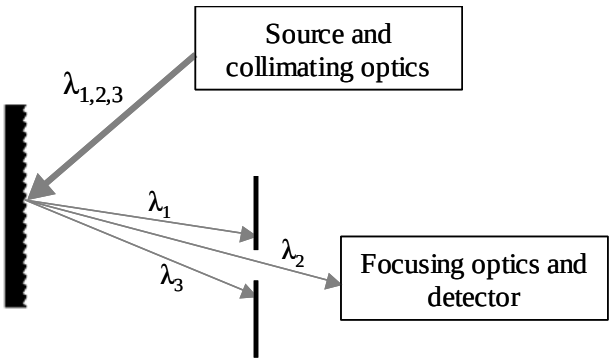


Figure 1.1 Schematics of a monochromator; a dispersive spectrometer. Narrow slits (an input and an output slit) are used to select a particular spectral element whose wavelength depends on the beam's incident angle on the grating. (Only the output slit is shown in this figure.)

1.2.2 Filter-based spectrometers

Filter-based spectrometers, or often simply called filter spectrometers, use one or more absorption or interference filters to transmit the selected range of wavelength, as illustrated in Fig. 1.2. As the beam passes through the filter, some of its spectral components are blocked through an absorption or interference process, while the desired spectral elements are transmitted. Various interference filters, from the ultraviolet through the far-infrared region, in various dimensions, are available as commercial-off-the-shelf items (e.g., Spectrogon AB, Taby, Sweden, and CVI Laser Corp., Albuquerque, NM, USA).

A commonly used spectroscopic configuration is that of a *filter-wheel* system, also available commercially. This system consists of a number of filters (with different wavelength responses) placed near the circumference of a rotating wheel. A spectral band is selected by positioning the wheel so that the beam falls on a particular filter. With this configuration, however, only a few discrete bands can be selected, rather than a continuous spectrum as with a monochromator. Another variation of the filter-based systems is the *tilting-filter* instrument.[1] In this instrument a spectral band is selected by changing the incident angle of the beam on the filter. However, the wavelength tuning range is rather limited at about ±3% of the center wavelength.

Because of the limited number of discrete wavelengths in filter-wheel instruments and the limited range of wavelength in tilting-filter instruments, filter-based spectrometers are dedicated to the specific analyses for which they are designed.

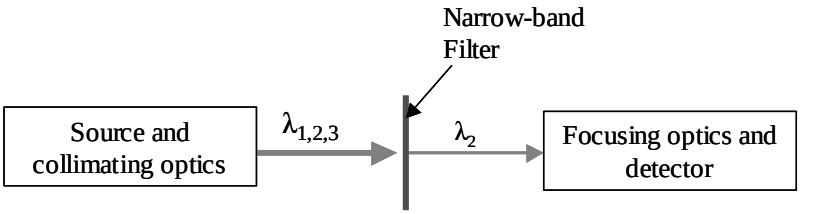


Figure 1.2 Schematic of a filter-based spectrometer. The filter selects the transmitted wavelength range through an absorption or interference process.

1.2.3 Fourier-transform spectrometers

A Fourier-transform (FT) spectrometer generates the spectrum of its source radiation by modulating the radiation in the time domain through interference, which is then Fourier transformed. The details of its operation are covered in Chapter 3. Here, a general overview of its operation is presented. Figure 1.3 shows the main components of an FT spectrometer and illustrates the overall process in obtaining the spectrum.

Briefly, the interference between the lights reflected off the moving and the stationary mirrors generates what is called an interferogram, *I(x)*, a modulated radiation signal as a function of the displacement of the moving mirror. This is the analog signal recorded at the photo detector, which encodes the wavelength or the wave number* information from the source radiation. A Fourier-transform routine is then performed on the interferogram to obtain the spectrum.

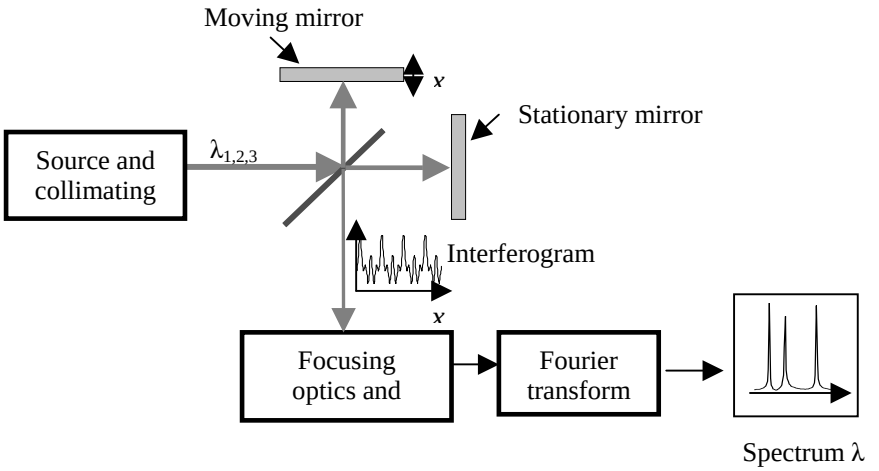


Figure 1.3 A Fourier-transform spectrometer using a Michelson interferometer. The interferogram contains the source's frequency information modulated in a time domain as a function of the moving mirror's displacement. The interferogram is Fourier-transformed to obtain the spectrum.

*Wave number = 1 / wavelength, given in cm^{-1}.

1.3 Advantages of FT Spectrometers

There are a number of acclaimed and widely publicized advantages of FT spectrometers when compared with their dispersive counterparts. However, only those advantages inherent to FT spectrometers' operating principles rather than particular engineering designs are discussed here. They are known as the *throughput (Jacquinot)* and the *multiplex (Felgett)* advantages.

Because filter instruments do not offer the kind of spectral range, spectral resolution, and versatility that FTS and grating instruments do they are not included in this comparison. It would be unfair to weigh the "higher-performing" instruments against the filter instruments.

The following discussions, although placed in the first chapter of the book, will better serve their purpose if the reader is already familiar with FTS operation. The throughput advantage discussions require familiarity with the finite spectral resolution imposed by the finite divergence angle of the beam entering the interferometer, covered in Sec. 3.3.2. Discussion of the multiplex advantage discussions requires knowledge of the spectrum generation from the interferogram illustrated throughout Chapter 3. Therefore, the reader who is unfamiliar with these subjects is advised to read Chapter 3 before continuing.

1.3.1 Throughput or Jacquinot advantage

Throughput, Θ, is defined as

$$\Theta = A\Omega\,, \tag{1.1}$$

where A is the area of the limiting aperture and Ω is the solid angle formed by the collimating or the focusing optics, given by

$$\Omega = \frac{\pi r^2}{f^2}\,, \tag{1.2}$$

where r is the radius of the illuminated optics and f is its focal length for a collimating system, as illustrated in Fig. 1.4. A solid angle is given in units of *steradians (sr)* and so throughput is usually given in units of $sr{\cdot}cm^2$. Essentially, throughput defines the allowable energy per unit time that the system can let through.

In spectrometers, a system's throughput can be limited either by the collimating optics in order to meet the resolution requirement (see Sec. 3.3.2) or by the detection optics (see Sec. 9.1). In medium- to high-resolution applications, throughput is generally limited by the collimating optics. If the optical throughput is indeed limited by the interferometer's collimating optics, then compare the maximum throughput allowable in FT spectrometers and monochromators to meet a spectral resolving power of R.

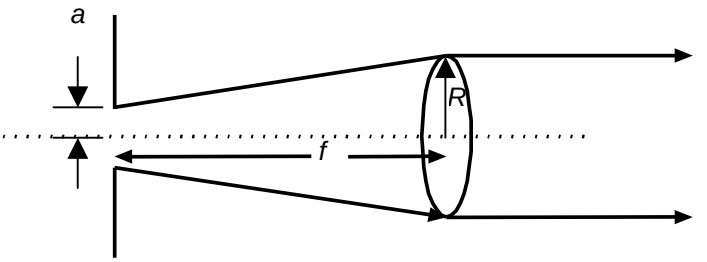


Figure 1.4 Optical throughput determined by an aperture.

First, the spectral resolving power R is defined to be

$$R = \frac{\upsilon_{\max}}{\Delta\upsilon}, \tag{1.3}$$

where $\upsilon_{\max}$ is the maximum wave number analyzed and $\Delta\upsilon$ is the spectral resolution. The following discussions are based on the analysis and results presented in Sec. 3.3.2.

For an FT spectrometer, by combining Eqs. (3.35) and (3.36) the resolving power R is limited by the angle of divergence α, and thus limited by the aperture:

$$R \approx \frac{1}{\alpha^2} \approx \frac{f^2}{a^2}, \tag{1.4}$$

using small angle approximations, where a is the radius of the aperture. Therefore, by combining Eqs. (1.1), (1.2) and (1.4), the throughput of FT spectrometers is given by

$$\Theta_{\text{FTS}} = \frac{\pi^2 r^2}{R}, \tag{1.5}$$

where πr^2 is equal to the area of the beam. If the mirror is completely "filled," then Eq. (1.5) can be written as

$$\Theta_{\text{interferometer}} = \frac{\pi}{R} A_{\text{mirror}}, \tag{1.6}$$

where A_{mirror} is the projected area of the mirror.

In the case of a grating spectrometer the throughput to achieve a resolving power R is limited by its collimating optics and the slit area, and is given by

$$\Theta_{\text{grating}} = \frac{l}{fR} A_{\text{grating}}, \tag{1.7}$$

where l is the height of the slit, f is the focal length of the collimating optics and A_{grating} is the projected area of the grating.[2] Generally, l/f does not exceed ~1/20 in today's grating spectrometers. Thus, for the same resolving power and similar

instrument size, FT spectrometers can offer more than ~60 times ($20 \times \pi$) higher energy-gathering capability. This advantage is called the *Jacquinot* or the throughput advantage. In weak-signal measurements, where the detector noise is dominant, spectral signal-to-noise ratio (SNR) increases proportionally with throughput. Thus, the Jacquinot advantage is greatly beneficial in that type of measurement.

1.3.2 Multiplex or Felgett advantage

A monochromator records a small band at a time, whereas a spectrograph records all bands simultaneously. Suppose a spectrum over N spectral bands is to be recorded within a period of T. In the case of the monochromator, each band is observed within a maximum time of T/N because it needs to be assessed separately. If the noise is random and independent of the signal level, the signal-to-noise ratio would be proportional to $(T/N)^{1/2}$, which is the *signal-integration* time employed within each band. Note that the signal increases proportionally with the integration time, and random noise increases proportionally with the square root of the integration time (see Sec. 10.2).

A spectrograph, on the other hand, records all bands simultaneously. Thus, each band is observed for the whole period T. In this case, the SNR would be proportional to $(T)^{1/2}$. Therefore, the SNR improvement obtainable by using the spectrograph is given by

$$\frac{\mathrm{SNR}_{\text{spectrograph}}}{\mathrm{SNR}_{\text{monochromator}}} = \frac{T^{1/2}}{(T/N)^{1/2}} = N^{1/2}. \tag{1.8}$$

This is called the multiplex advantage.

Fourier-transform spectrometers have a similar advantage first described by Felgett.[3] Essentially, all bands or wavelength components in FT spectroscopy are observed simultaneously and recorded as the interferogram. This is because the interferogram essentially contains the *sum* of the interference patterns generated by each wavelength component.

Considering the same scenario, time T measures the source radiation and produces a spectrum of N spectral bands. In this case, however, T is used to record the interferogram. If the noise is random and independent of the signal, the signal-to-noise ratio of the interferogram, SNR^I, is given by

$$\mathrm{SNR}^I_{\mathrm{FTS}} \propto \left(\frac{T}{N_I}\right)^{1/2}, \tag{1.9}$$

where N_I is the number of interferogram points. In going from the interferogram to the spectrum, SNR is improved by $N_I^{1/2}$.[4] (This point is explained in more detail shortly). Therefore, the spectral SNR of an FTS, $\mathrm{SNR}^S_{\mathrm{FTS}}$, is

$$\mathrm{SNR}^{S}{}_{\mathrm{FTS}} = \mathrm{SNR}^{I}{}_{\mathrm{FTS}} \times N_I^{1/2} = \left(\frac{T}{N_I}\right)^{1/2} \times N_I^{1/2} = T^{1/2}. \tag{1.10}$$

The SNR improvement obtainable by using an FTS is thus

$$\frac{\mathrm{SNR}_{\mathrm{FTS}}}{\mathrm{SNR}_{\mathrm{monochromator}}} = \left(\frac{T}{T/N}\right)^{1/2} = N^{1/2}. \tag{1.11}$$

The relationship between the number of interferogram points N_I and the number of spectral elements N are given in Sec. 7.4. Sakai presents a thorough statistical analysis[4] that explains the spectral SNR improvement as a function of the interferogram points as stated in Eq. (1.10). Here, the analytical claim is verified through a numerical analysis and the result is presented in Fig. 1.5. In the simulation, an interferogram of a monochromatic signal is generated upon which a random noise is added. This noise-added interferogram is then Fourier transformed to obtain the spectrum. The SNR of the interferogram is obtained by taking the ratio of the mean value of the interferogram and the standard deviation of the noise. Regardless of the number of interferogram points, it results in the same value. The SNR value of the spectrum is obtained by measuring the ratio of the spectral peak and the standard deviation of the spectral noise.

This result may be somewhat counter-intuitive because it suggests that for a certain spectral range analyzed, the spectral SNR increases as the resolution increases. As discussed in Chapter 6, the spectral resolution is increased by scanning the moving mirror to a greater distance (thus collecting more interferogram points). By doing this, the spectral SNR increases since the interferogram contains the sum of all wavelength components. Thus, the more data collected, the stronger the signal at each band. Figure 1.6 illustrates this by comparing the signal and the noise spectra of a monochromatic signal produced

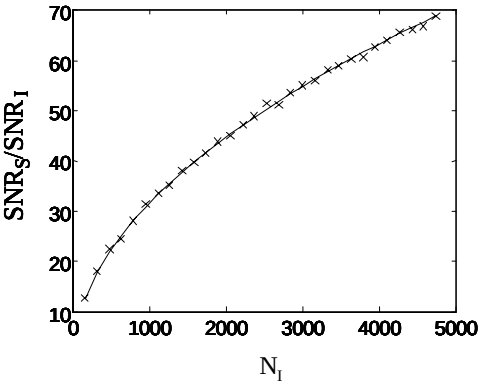


Figure 1.5 Numerical simulation results showing the increase of spectral SNR as a function of the number of interferogram points N_I. Solid line: theoretical line of $N_I^{1/2}$. X line: results of the numerical simulation.

by Fourier-transforming interferograms with two different quantities of points (different spectral resolutions). As seen in the figure, the magnitude of the spectral noise increases by about a factor of 2, while the magnitude of the peak spectrum increases by a factor of 4. Thus, in this case the SNR improvement is a factor of 2.

1.4 Discussion on FT Spectrometer Advantages

The actual SNR improvements obtained by an FT spectrometer depend on various other instrumental and procedural factors. In some measurements, none of the advantages of FT spectrometers may really apply. The various sources of noise and errors that may limit the achievable spectral SNR obtainable by FT instruments is discussed later. It is one of the goals of this book to equip the reader with enough knowledge about the design and the instrumentation aspects of FT spectrometers to be able to critically evaluate their tangible benefits for the actual applications.

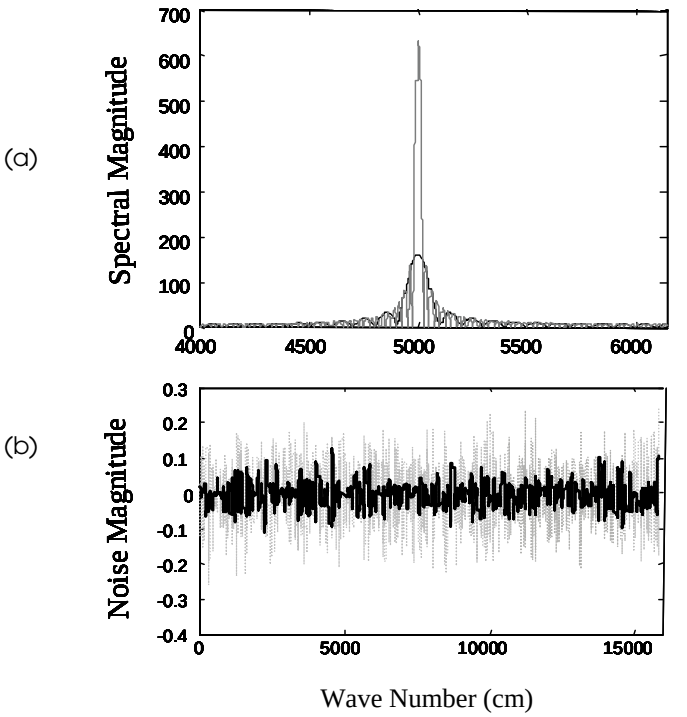


Figure 1.6 (a) transmission spectra of a monochromatic signal taken with two different sizes of interferogram data points. The spectrum in gray has four times the interferogram points than the spectrum in black. (b) Spectral noise of the two spectra shown in (a). It can be seen that where the noise magnitude increases by a factor of 2 (b), the signal magnitude increases by a factor of 4 (a), indicating an SNR increase by a factor of 2.

As a brief illustration, consider an application in the visible and/or the near-infrared range (~400 to ~2500 nm). For measurements in such a high wave number (short wavelength) radiation, the FT spectrometer requires a high-precision mirror drive and an extremely accurate interferogram data sampling, which will be discussed in Chapters 4 and 6, respectively. Otherwise, the resulting spectral SNR is very low. If spectral SNR is indeed limited by the quality of the mirror drive or the data sampling accuracy, none of the FT spectrometer's advantages previously discussed applies. Thus, in order to improve the SNR, better motion- and data-clocking systems should be used. In this case, one needs to carefully assess the benefits of using an expensive, "high-performance" FT spectrometer compared with a less expensive, grating-based instrument. The deciding factor may be the resolution required and the wavelength range to be measured at a given time, which translates to the number of spectral elements. With a large number of spectral elements, the multiplex advantage may be necessary or highly beneficial. The throughput advantage is also relevant if the radiation source is weak.

References

1. J. Scheer, "Programmable tilting filter spectrometer for studying gravity waves in the upper atmosphere," *Applied Optics*, **26**(15), pp. 3077–3082 (1987).
2. E. V. Loewenstein, "Fourier spectroscopy: an introduction," *Proceedings of Aspen International Conference on Fourier Spectroscopy*, pp. 3–17 (1970).
3. P. Felgett, "A propos del la théorie du spectrum etre interferential multiplex," *le Journal de Physique et le Radium*, **19**, pp. 187 (1958).
4. H. Sakai, "Consideration of the signal-to-noise ratio in Fourier spectroscopy," *Proceedings of Aspen International Conference on Fourier Spectroscopy*, pp. 19–41, (1970).

Chapter 2
Signal-to-Noise Ratio

To quantify the instrument's performance, the spectral *signal-to-noise ratio* (SNR) is used as the main measure throughout this book. The term has been used somewhat inconsistently; in some cases, it is used to quantify spectral repeatability, and in others, it is used to quantify spectral accuracy. Thus, it is appropriate to start with a clear definition of the term as it applies throughout this book.

2.1 Signal-to-Noise Ratio Defined

In this book, the SNR measures the instrument's ability to reproduce the spectrum from the same sample, the same conditions, and the same instrumental configurations over a certain amount of time. This, in fact, is a measure of spectral repeatability, which measures the ability of the instrument to detect certain changes in the spectrum such as those caused by changes in the sample's spectral characteristics. Therefore, noise is the measure of the spectral deviations between measurements, regardless of the output spectrum's proximity to the "true" value.

Accuracy, on the other hand, is a measure of the discrepancy between the actual measured value and the "true" standard or calibrated value. Wavelength accuracy of a spectrometer is crucial. It is important for the instrument to be able to conform to the "calibrated standard" in producing wavelength information within the instrument's intended resolution. For example, if the true absorption peak of molecule x is 4000.2 nm, the instrument should be able to give the correct peak wavelength information within its designed resolution. Thus, following the example, if the instrument has a designed resolution of 1 nm, then it should register the peak of molecule x at 4000 nm. This is required for proper information transfer between instruments.

The spectral magnitude accuracy of FT spectrometers, however, is meaningless because it depends on various factors that are difficult to "standardize," such as the optics' transmission properties, the detector's characteristics, the speed of the moving mirror, the electronics' bandwidth, among various other factors. However, this does not generally cause a problem

because the instrument's transfer function is usually *zeroed-out* by first taking the measurement of a reference spectrum. The reference spectrum may simply be that of the ambient air, or a certain reference sample.

Fourier-transform spectroscopic measurements generally involve two steps: first is the recording of the reference spectrum, and second is the recording of the sample spectrum. This should apply to both the absorption and the emission studies. Theoretically, the results are then independent of the instrument's transfer function, and are transferable between different instruments. For this to be true, however, the recorded intensity at each wavelength has to have high repeatability, and be within the linear range of the detection unit. The importance of the detector's linearity is discussed in Sec. 8.4.4.

2.2 Quantifying Signal-to-Noise Ratio

Most of the measurements using FT spectrometers rely on the single beam technique that involves taking the ratio of the sample's transmission spectrum to the reference spectrum at two separate times. Consequently, spectral noise $N(\upsilon)$ can be computed according to the following equation:

$$N(\upsilon)=1-\frac{T_a(\upsilon)}{T_b(\upsilon)}, \tag{2.1}$$

where T_a and T_b are the transmission spectra of the same sample/buffer taken at two different times, as illustrated in Fig. 2.1. This method can also be applied to the double-beam measurements, in which case T_a and T_b are taken simultaneously. In the case of zero noise, $N(\upsilon)$ will be a flat line at zero across the wavelength range. A *100% line* is defined by 100 × $T_a(\upsilon)$ / $T_b(\upsilon)$ and is usually specified in the FT spectrometer's manufacturer's data sheet.

The *root-mean-square* (rms) value of the spectral noise N_{rms} can then be computed as

$$N_{\text{rms}}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[N(\upsilon_i)\right]^2}, \tag{2.2}$$

where n is the number of spectral elements being observed. The range of the spectral elements should cover only the spectral range of interest. Note that the rms value is usually three to five times smaller than the peak-to-peak noise value. Finally, spectral SNR can be computed as

$$\text{SNR}=\frac{1}{N_{\text{rms}}}. \tag{2.3}$$

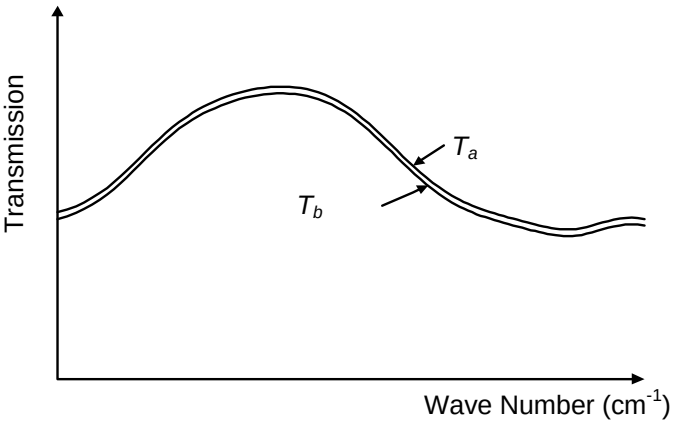


Figure 2.1 An illustration of the procedure to obtain an experimental SNR value. Two spectra of the same sample.

Figure 2.2 shows a theoretical illustration of computed spectral noise. In quantitative applications where sample concentration is deduced from the change in spectral magnitude, it is useful to compare the spectral noise with the actual absorption spectrum of the sample. From such a graph, it can be deduced whether the instrument is capable of actually resolving the spectral feature(s) of the absorbing sample. For example, the measurement is a sample concentration that corresponds to 10^{-4} AU (absorbance unit*), the spectral noise within the wavelength band region has to be significantly less than 10^{-4} AU.

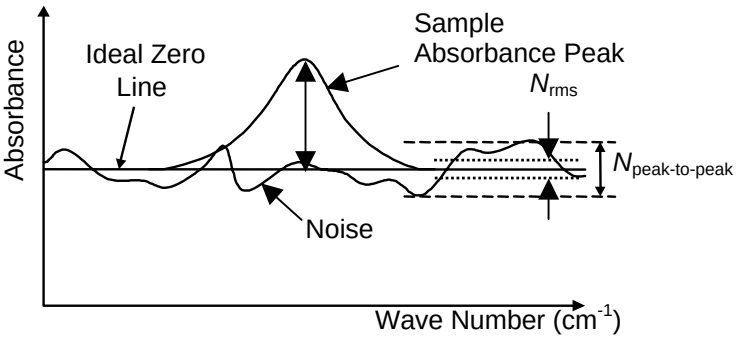


Figure 2.2 Illustration of spectral noise and its comparison with an absorption spectrum. Spectral noise is computed by taking the ratio of the two single-beam spectra, as shown in Fig. 2.1.

* Absorbance is defined as

$$A(\upsilon_i) = -\log_{10}\left(\frac{T_{\text{sample}}(\upsilon_i)}{T_{\text{reference}}(\upsilon_i)}\right) AU \ .$$

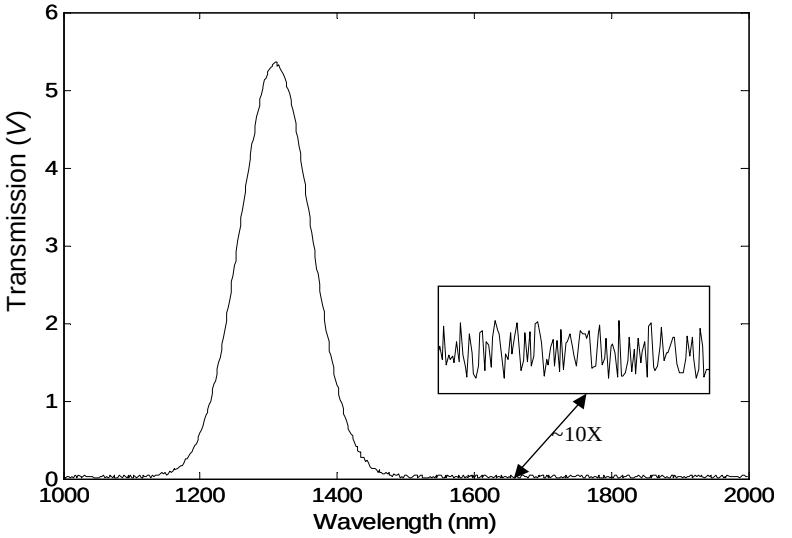


Figure 2.3 Spectral SNR calculation using the transmission spectrum.

One might wonder how this quantification method relates to the more "conventional" way of measuring the SNR, that is by comparing the peak of a transmission spectrum with the noise (which is usually easier to see at the base of the transmission spectrum), as illustrated in Fig. 2.3. After all, in its simplest form the SNR is

$$\mathrm{SNR} = \frac{\mathrm{signal}_{\mathrm{received}}}{\mathrm{noise}}. \tag{2.4}$$

The two SNR quantification methods do in fact produce similar SNR values. With the "conventional" method the SNR is calculated as

$$\mathrm{SNR} = \frac{\mu(\upsilon)}{\mathrm{rms}\left[T(\upsilon) - \mu(\upsilon)\right]}, \tag{2.5}$$

where μ is the ideal mean or the "noiseless" transmission spectrum and T is the measured spectrum. It follows that

$$\mathrm{SNR} = \frac{\mu(\upsilon)}{\mu(\upsilon)\mathrm{rms}\left(\frac{T(\upsilon)}{\mu(\upsilon)} - 1\right)} = \frac{1}{\mathrm{rms}\left(\frac{T(\upsilon)}{\mu(\upsilon)} - 1\right)}, \tag{2.6}$$

where rms [g(υ)] is the rms value of the function g(υ) as given by the Eq. (2.2). This result is similar to that shown by Eq. (2.3). The crucial difference of note is that with this method, it is assumed that the noise spectrum can be distinguished from the real spectrum. This is relatively easy to do if the noise has much higher

"frequency" features than the signal, as is the case for the spectrum depicted in Fig. 2.3. However, if the noise is slow varying or has a similar frequency to the signal, then it would be more difficult to distinguish.

2.3 Practical Considerations

The SNR value depends on many experimental factors as well as instrumental configurations, including the averaging and integrating time, the radiation power that falls on the detector, the spectral range observed, the diameter of the beam entering the interferometer, etc., as is demonstrated throughout this book. Thus, one needs to be aware of and take into account these conditions when reporting the SNR value of an instrument or when evaluating the manufacturer's specified SNR.

It is highly recommended that one perform an experimental SNR measurement under the same conditions or instrumental configurations as those to be used in the real measurement. Consider as an example, a near-infrared instrument used to measure an analyte in a water solution. Measuring the SNR by taking single-beam spectra of the background air (to obtain the 100% line) is not appropriate since the transmitted spectral range, the transmitted power radiation, and the aperture setting most likely differ from those used in the actual measurements of the analyte. Evaluating the SNR with use of a controlled, aqueous solution or a set of appropriate optical filters to replicate the real conditions produces a more meaningful result.

Furthermore, the SNR calculation procedure given in Sec. 2.2 should not be applied blindly. Sometimes, a slight adjustment or modification to the procedure or the equations can make a better "fit" in certain situations. For example, Fig. 2.4 shows a noise spectrum produced by direct application of Eq. (2.1). If the slow-varying baseline variation does not affect the sample *detectability*, it should be removed before applying Eq. (2.3) to obtain the SNR. Otherwise, the computed SNR value may underestimate the actual detectability by several orders of magnitude.

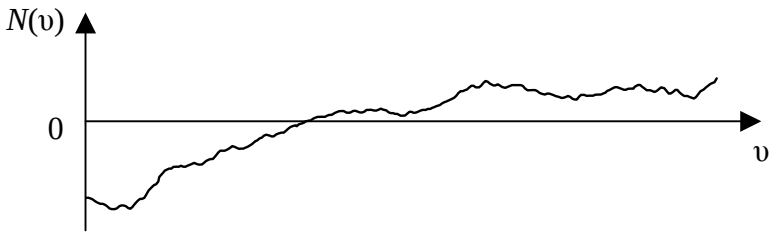


Figure 2.4 Noise spectrum showing a slow-varying baseline feature.

Another practical matter to consider is the wavelength region to include in the spectral SNR calculation. FT spectrometers produce spectra with a wave number range between zero and υ_{max}, with υ_{max} determined by the sampling period. The wavelength range for the SNR calculation, however, should include only the range of interest that depends on the sample's spectral response. Including the entire range underestimates the achievable SNR of the actual measurement.

Chapter 3
Principles of Interferometer Operation

The interferometer is the core of every Fourier-transform spectrometer. Today's FT spectrometers use a variety of interferometer designs. However, they are all still based on the simple, yet historically most important, Michelson interferometer.

In this chapter, the operating principle of the Michelson interferometer for FT spectroscopy is discussed. It is the objective of this chapter to provide a thorough physical understanding of how a spectrum is generated in an FT spectrometer. Enough mathematics is used to aid comprehension. First, a qualitative overview is provided, which is followed by a more detailed explanation starting with the wave description of light. Then, the factors that limit the output spectral resolution are explored, and finally, the interferogram processing techniques sometimes necessary to obtain accurate spectra are briefly discussed.

3.1 Overview

Figure 3.1 shows a schematic of a Michelson's interferometer. It consists of a beamsplitter and two plane mirrors that are perpendicular to each other. One of the plane mirrors, *M*2, moves linearly in the direction shown by the arrow. As light enters the interferometer, it is amplitude divided at the beamsplitter. Approximately one-half of the light is transmitted and the other half is reflected. The transmitted and reflected beams are then reflected at mirrors *M*2 and *M*1, respectively. The beams are then recombined at the beamsplitter and detected by a photodetector.

Under first consideration is light from a monochromatic source. When L_1 is equal to L_2, the two beams travel the same distance from the point they leave the beamsplitter to the point where they recombine at the beamsplitter. Being in-phase, they interfere constructively (Fig. 3.2) and the detector sees a maximum intensity. As L_2 moves away from the zero *optical path difference* (OPD), the

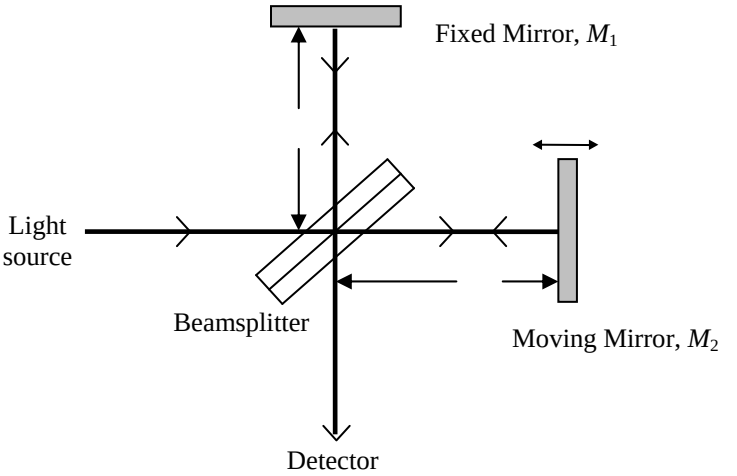


Figure 3.1 Schematic of a two-plane mirror Michelson interferometer.

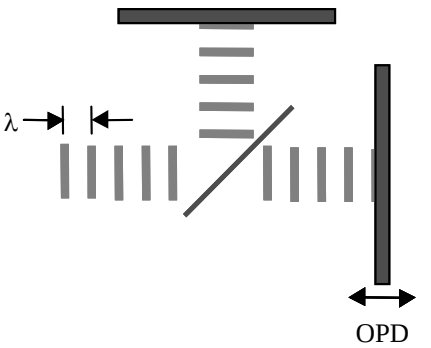


Figure 3.2 Monochromatic waves at zero OPD.

intensity starts to decrease as phase difference is introduced. Note that the OPD is equal to twice the mirror retardation distance because the OPD corresponds to the distance traveled by the beam to and from the moving mirror. When the OPD is equal to λ/2, destructive interference occurs as the two recombined beams become out-of-phase with each other. Thus, as the moving mirror travels at a constant velocity, the detector sees a sinusoidal-varying intensity. This sinusoidal signal is a function of mirror *M*2 displacement with a period of λ/2. The recombined interfering beams' intensity fluctuation as a function of mirror displacement is called an interferogram. The interferogram is then Fourier transformed to obtain the spectrum. Figure 3.3 shows an interferogram from a monochromatic source and its spectrum.

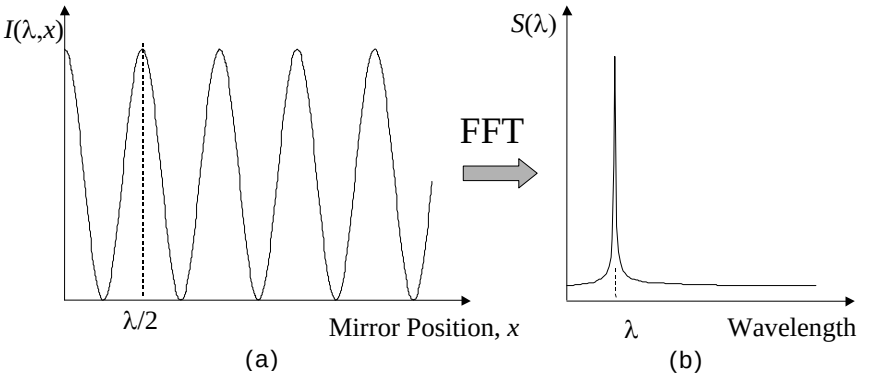


Figure 3.3 (a) Interferogram of monochromatic light; (b) its spectrum. "FFT" indicates fast Fourier-transform operation.

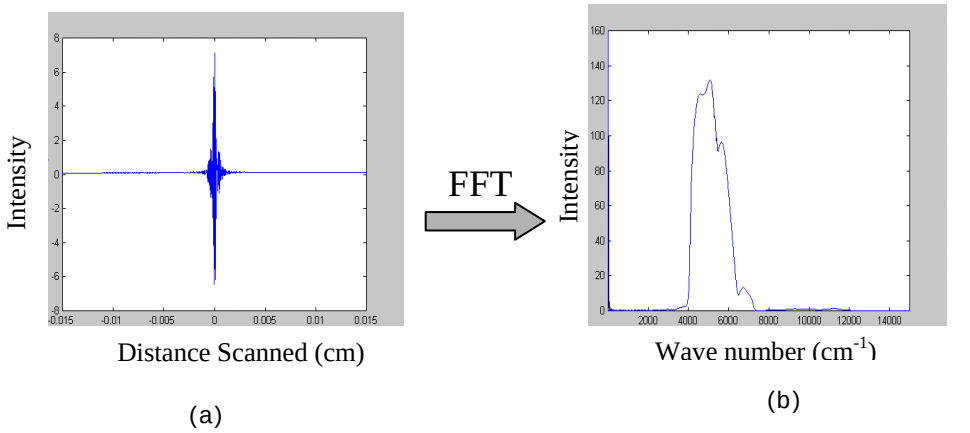


Figure 3.4 (a) Interferogram of polychromatic light; (b) its spectrum. "FFT" indicates fast Fourier-transform operation.

When the radiation comes from a broadband source, the interferogram contains a peak at zero mirror retardation, and when all wavelength components interfere constructively, they decay quickly as the mirror moves. Figure 3.4 shows an experimentally obtained interferogram of a broadband near-infrared source and its resulting spectrum.

3.2 Quantitative Explanation

3.2.1 Light as a wave

To fully understand light interference and interferogram generation, light as an electromagnetic wave must be considered. A monochromatic light contains a single-frequency wave, though no real waves are truly monochromatic. A source

that contains a narrow enough band of frequencies and is stable for a long enough period is a practical definition of monochromatic light. For this discussion, sources such as helium-neon lasers that emit a single spectral peak are considered monochromatic. A traveling monochromatic wave with speed v in the positive z-direction can be represented mathematically by

$$U(z,t) = A\sin k\left(z - vt\right), \tag{3.1}$$

where A is the wave amplitude, t is time, and k is a positive constant known as the propagation number

$$k = \frac{2\pi}{\lambda}. \tag{3.2}$$

Figure 3.5(a) shows the temporal magnitude fluctuation of the monochromatic wave (as a function of time), with frequency f. Figure 3.5(b) shows the wavefronts at a snapshot of time. Each wavefront is separated from the next by λ. The frequency f and the wavelength λ of the wave are related by

$$\lambda = \frac{v}{f}, \tag{3.3}$$

where

$$v = \frac{c}{n}, \tag{3.4}$$

where c is the speed of light in a vacuum, ~3×10^8 m/s, and n is the refractive index of the medium.

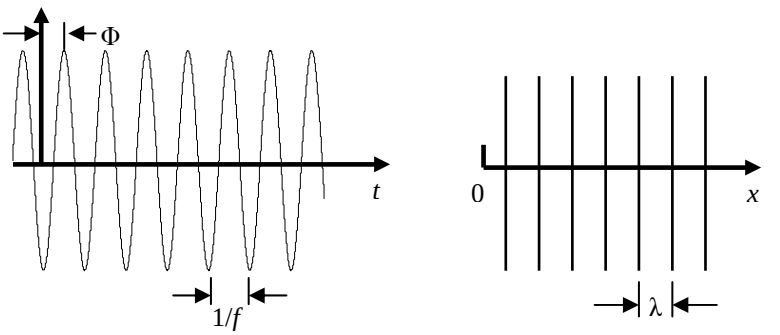


Figure 3.5 Temporal and spatial representations of a plane monochromatic wave traveling in the z direction. The phase with respect to a fixed reference of time is indicated by "Φ."

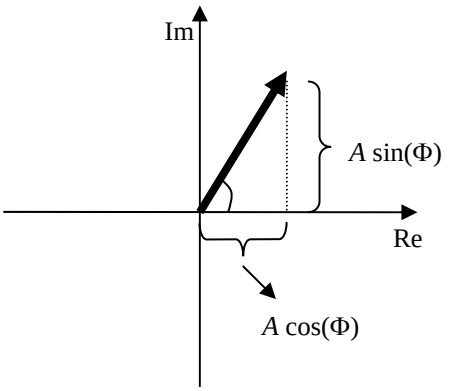


Figure 3.6 An Argand diagram used to represent a complex number in terms of its real and imaginary components.

For interference analysis, it is beneficial to describe the wave with a complex number representation. The monochromatic wave in Eq. (3.1) can be represented by the complex function

$$U = A\left(\cos\Phi + j\sin\Phi\right) = Ae^{j\Phi}, \tag{3.5}$$

where A is an amplitude constant, $j = \sqrt{-1}$, and

$$\Phi = kz - kvt. \tag{3.6}$$

It is also helpful to picture the complex number as a vector-like entity consisting of its real and imaginary components, called an *Argand* diagram, as shown in Fig. 3.6. The arrow and its phase angle constitute a *phasor*.

3.2.2 Measurable light quantity

Eventually, the radiation power is converted into an electrical signal at the photodetector that can then be processed, digitized, and stored. The question now is, what does the detector measure? After all, if the wave's electric or magnetic field temporal variations could be measured, there would be no need to use interferometry to create modulation through interference in the first place. It is possible to have obtained its spectrum simply by Fourier transforming the recorded signal as a function of time. However, no current detectors exist that have a high enough bandwidth to detect the high frequency of light (1000-nm light has a temporal frequency of ~3×10^{14} Hz).

Rather, the detected signal is the average of the optical intensity I integrated over a certain period determined by the bandwidth of the detector. Intensity is the optical power per unit area that is equal to the absolute square of the light's complex amplitude;[1] thus, the intensity of the detected signal is given by

$$I = \left\langle |U| \right\rangle^2. \tag{3.7}$$

The angled brackets ⟨⟩ represent an averaging process over a longer period than that of the light wave.

3.2.3 Interference and superposition

In the Michelson interferometer, the two electromagnetic waves recombine at the beamsplitter and interfere. The relative *optical path length* (OPL) of the two beams determines the intensity of the recombined light.

First, consider interference between two monochromatic waves of the same frequency. When two monochromatic waves of complex amplitudes U_1 and U_2 of the same frequency f_o are summed, the result is a monochromatic wave of frequency f_o and a complex amplitude:

$$U = U_1 + U_2 . \tag{3.8}$$

The *phasor addition* method is used to obtain the mathematical expression of the interference-modulated signal. This method involves summing the phasors, as one would sum vectors, and as illustrated in Fig 3.7. Thus, the sum of the two waves is calculated to be

$$U^2 = \left[\mathrm{Re}(U_1) + \mathrm{Re}(U_2)\right]^2 + \left[\mathrm{Im}(U_1) + \mathrm{Im}(U_2)\right]^2 , \tag{3.9}$$

where $\mathrm{Re}(U_1)$ and $\mathrm{Im}(U_1)$ are the real and the imaginary part of the wave equation, respectively. Working through the algebra,

$$U^2 = U_1^2 + U_2^2 + 2(U_1 U_2)^{1/2} \cos(\Phi_2 - \Phi_1) \tag{3.10}$$

is obtained. Given Eq. (3.7), the detected light intensity can be calculated to be

$$I = I_1 + I_2 + 2(I_1 I_2)^{1/2} \cos(\Phi_2 - \Phi_1) . \tag{3.11}$$

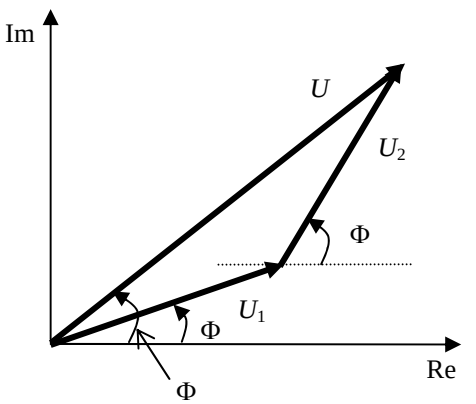


Figure 3.7 Phasor addition method; summation of two phasors.

Equation (3.11) is usually called an interference function, and when rearranged

$$I = (I_1 + I_2)\left[1 + \frac{2(I_1 I_2)^{1/2}}{I_1 + I_2}\cos\delta\right], \tag{3.12}$$

where δ is the relative phase between waves 1 and 2.

The next task is to relate this result to interference at the Michelson interferometer. First, to what the relative phase δ corresponds must be determined. Consider the two monochromatic beams that are being recombined at the beamsplitter. The relative distance, or the OPD, that the two beams travel is

$$\text{OPD} = 2x. \tag{3.13}$$

Constructive interference occurs at

$$\text{OPD} = n\lambda, \tag{3.14}$$

where n = 1, 2, 3, etc. Thus, it can be seen that the relative phase between the two interfering waves, in units of radians, is given by

$$\delta = 2\pi\frac{\text{OPD}}{\lambda} = \frac{4\pi x}{\lambda}, \tag{3.15}$$

as illustrated in Fig. 3.8.

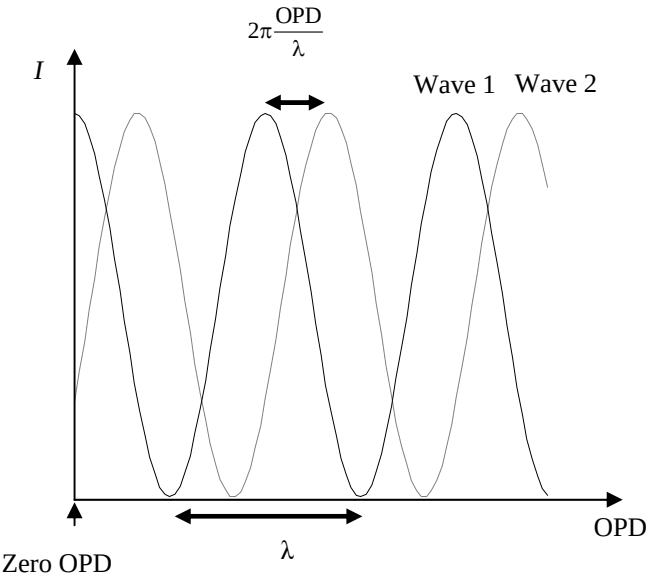


Figure 3.8 Relationship of the relative phase δ in the general interference function with the OPD in the Michelson interferometer.

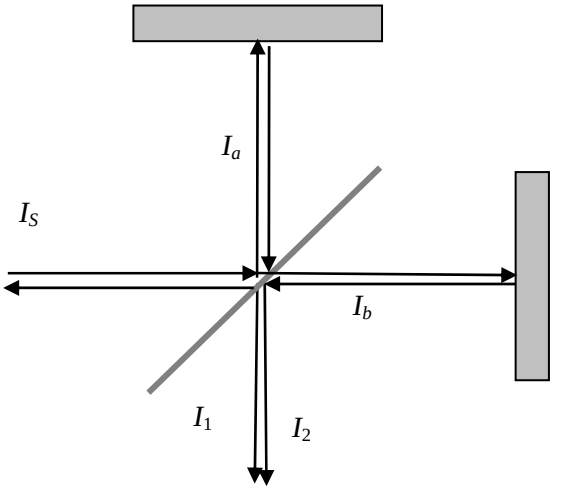


Figure 3.9 Beam intensities at various "arms" of the interferometer.

For an ideal beamsplitter, the percentage of transmittance (%T) is equal to the percentage of reflectance (%R), and is equal to 50%. An ideal mirror has 100% reflectance. Assuming an ideal beamsplitter and ideal mirrors,

$$I_a = I_b = 0.5 I_S , \tag{3.16}$$

$$I_1 = 0.5 I_a , \tag{3.17}$$

and

$$I_2 = 0.5 I_b . \tag{3.18}$$

where $I_a, I_b, I_1, I_2,$ and I_S are shown in Fig. 3.9.

Thus, the interference function stated in Eq. (3.12) for a Michelson interferometer with ideal optics becomes

$$I(x) = 0.5 I_S \left[1 + \cos\left(\frac{4\pi x}{\lambda} \right) \right]. \tag{3.19}$$

In most cases, only the ac-modulated part of the interferogram is of interest. In fact, "interferogram" often refers to only the ac-modulated part of Eq. (3.19),

$$I_{ac}(x) = 0.5 I_S \cos\left(\frac{4\pi x}{\lambda} \right). \tag{3.20}$$

3.2.4 Polychromatic source

Polychromatic or broadband radiation refers to radiation that contains a wide range of frequencies. It is known that any spectrum can be synthesized by a

superposition of sinusoidal waves. In the discrete case, it is illustrated by the following equation:

$$f(y) = C_0 + C_1 \cos\left(\frac{2\pi}{\lambda_1} y\right) + C_2 \cos\left(\frac{2\pi}{\lambda_2} y\right) + \ldots, \tag{3.21}$$

where the *C*s are constants. Thus, the interference and superposition of the monochromatic waves as discussed in the previous sections still apply in the polychromatic case. The difference is just that now, in essence, there is a number of monochromatic sources. Thus, the interference function (or the interferogram) is now given by

$$I_{ac}(x) = \sum_{\upsilon=\upsilon_{\min}}^{\upsilon_{\max}} C(\upsilon)\cos(4\pi\upsilon x), \tag{3.22}$$

where $C(\upsilon)$ denotes the spectral distribution of the polychromatic radiation (intensity of each particular wave number υ) recorded by the detector.

An interferogram of a polychromatic source is characterized by a large signal at zero mirror retardation, which quickly decays to a mean value. The large signal at zero-mirror retardation or zero OPD is usually called the centerburst. Figure 3.4 depicts an example of an interferogram from a polychromatic source.

3.2.5 Fourier-transform routine

Now that the interferogram has been obtained, it is time to extract the spectral distribution information, $C(\upsilon)$. This is done by taking the Fourier transform of the interferogram. It can be shown that

$$C(\upsilon) = \int_{-\infty}^{\infty} I(x)\cos(4\pi\upsilon x)dx. \tag{3.23}$$

This operation is called *cosine Fourier transformation*. The mathematical details of the Fourier-transform routine are out of the scope of this book. Interested readers can find such details in numerous mathematics textbooks.

In practice, a cosine Fourier transformation alone is not enough to accurately obtain the spectrum from the interferogram. This is because, rather than being symmetrical about the zero OPD (as cosine waves are), the real interferograms contain phase errors that may be a complex function of the retardation distance, making them asymmetrical. Thus, the complex Fourier transform, containing both the sine and cosine components, is usually used. It is represented by

$$C(\upsilon) = \int_{-\infty}^{\infty} I(x)\exp(-i4\pi\upsilon x)dx, \tag{3.24}$$

where

$$\exp(i\omega) = \cos\omega + i\sin\omega. \tag{3.25}$$

Of course, this operation is most commonly done using the Fourier transform digital method of Cooley and Tukey.[2] This method is referred to as *fast Fourier transform* (FFT), which reduces the amount of computation steps (thus time) significantly. In order to use this technique, however, the number of data points must be at a power of two. This is generally not a problem in FT spectroscopy. Null data points (discussed later in this chapter) can be added to the original interferogram data to meet this requirement.

3.3 Theoretical Resolution

The spectral resolution of a spectrometer, $\Delta\upsilon$, is a measure of its ability to qualitatively distinguish two spectral peaks that are very close to each other. There are a few definitions that describe two spectra being resolved such as the *baseline criterion* and the *Rayleigh* criterion;[3] however, they are defined rather arbitrarily.

A commonly used figure, *spectral resolving power*, is given by

$$R = \frac{\upsilon_{\max}}{\Delta\upsilon}, \tag{3.26}$$

where $\upsilon_{\max}$ is the maximum wave number (shortest wavelength) for which the spectrometer is designed to operate.

Fundamentally, the resolving power of an FT spectrometer is limited by two conditions: the maximum OPD, and the beam's angle of divergence. Thus, the two variables necessary to meet a certain resolution requirement are the mirror-scan distance and the aperture size of the interferometer's collimating optics. In practice, errors from other sources, such as mirror tilt, may degrade the output resolution further.

3.3.1 Retardation distance

The range of motion of the moving mirror is directly related to the maximum achievable resolution, as given by the following relationship:

$$\Delta\upsilon = \frac{1}{\mathrm{OPD}}. \tag{3.27}$$

Readers who are interested in its thorough mathematical derivation are referred to Bell.[4] Here, a qualitative argument is presented, and its accuracy verified through a numerical simulation.

Consider two monochromatic waves of slightly different wave numbers, υ_1 and υ_2, entering an interferometer. Since the wavelengths are very close to each other, it may take the two waves many cycles before they become out of phase, and similarly, back in phase. Starting at zero OPD, the two waves become in phase at an OPD equal to $1/\Delta\upsilon$, as illustrated in Fig. 3.10(a). The OPD must

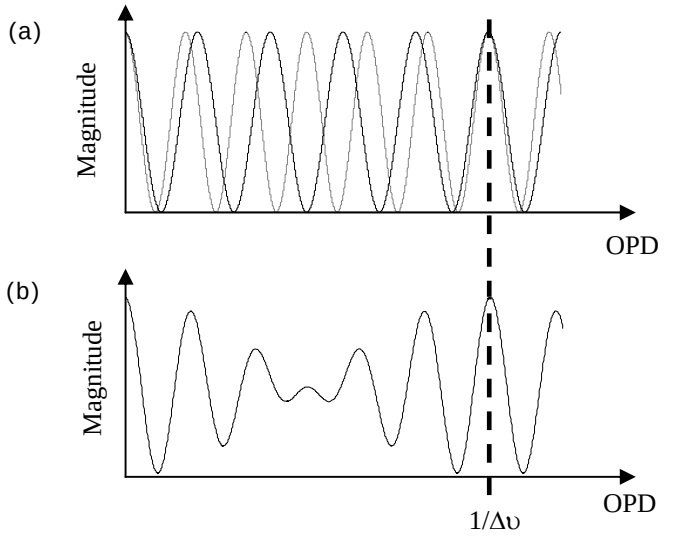


Figure 3.10 Resolution as a function of OPD. (a) Two monochromatic waves. Solid line: υ_1; dashed line: υ_2; $\Delta\upsilon = |\upsilon_1 - \upsilon_2|$. (b) The resulting interferogram (sum of the two waves).

exceed $1/\Delta\upsilon$ and let one full cycle of the envelope be recorded before the interferogram contains the complete information to distinguish the two waves. Also, consider the resolution limitation of an FT spectrometer by looking at its *instrumental line shape* (ILS). The ILS can be pictured as the resulting output spectrum when a purely monochromatic signal (having an infinitely small spectral width) is the input radiation. The origin of the ILS can be understood by considering the Fourier integral given in Eq. (3.24). The "ideal" integration limits, as noted in the equation, are negative and positive infinity. This means that the mirror needs to travel an infinite distance, which of course is not possible. Putting finite limits to the integral results in a *sinc* function, as illustrated in Fig. 3.11. This function is the ILS of a Fourier transform spectrometer without an *apodization* function (or can be considered as using a *top-hat* apodization function). Apodization is discussed in more detail in Sec. 3.4.1.

It turns out that the width of the sinc function is inversely proportional to the width of the top-hat function between the zero-line crossings, as shown in Fig. 3.16. This means that the optical path difference is approximately related to the spectral resolution by

$$\mathrm{OPD} \approx \frac{1}{\Delta\upsilon}, \tag{3.28}$$

which agrees with Eq. (3.27).

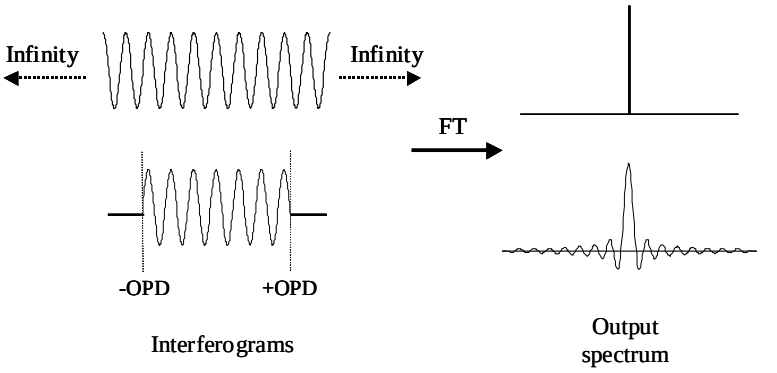


Figure 3.11 Instrument line shape (ILS) due to finite optical path difference.

Now the accuracy of the analytical investigations is checked through a numerical simulation. In the simulation, the following parameters are used:

$\upsilon_1 = 2000\ \text{cm}^{-1}$

$\upsilon_2 = 2005\ \text{cm}^{-1}$

Sampling period: 158.2 nm

Zero filling: an equal number of points on both sides of the interferogram tails are added to make 2^{18} total number of interferogram points.

How far does the moving mirror have to go before the spectral bands are resolved when the Fourier transformation takes place? Figure 3.12 shows two output spectra resulting from two different OPDs. It is shown that the two bands are fully resolved when the OPD is equal to 0.2 cm, consistent with Eq. (3.27).

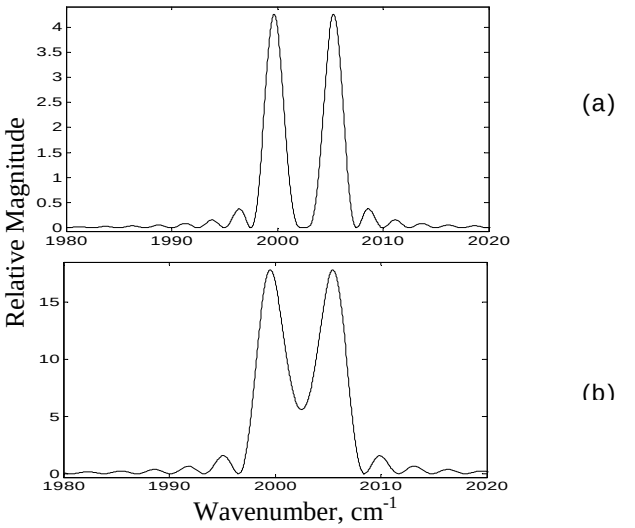


Figure 3.12 (a) Output spectrum of two close bands with OPD = 0.2 cm (OPD = $1/\Delta\nu$). (b) Output spectrum of two close bands with OPD = 0.16 cm (OPD < $1/\Delta\nu$).

3.3.2 Divergence angle

So far, it has been assumed that the wave phases of the plane wave (a perfectly collimated beam) do not have spatial dependence. In reality, the beam entering the interferometer has a finite divergence angle, which for incoherent sources can be large. The divergence angle of the beam entering the interferometer has a direct consequence to the achievable resolution, which will be seen shortly. Thus, collimating optics is used to limit the angle of divergence, which can simply consist of an aperture and a positive lens, as depicted in Fig. 3.13.

In this figure, two extreme rays are shown undergoing different paths due to the finite divergence angle that is limited by an aperture with radius *a*. This divergence-limiting aperture is often called the *Jacquinot stop*. To easilyunderstand the effect of divergence on resolution, again consider a monochromatic wave. As seen earlier in this chapter, the wavelength information of the light source is obtained from the frequency or the period of the interferogram, which in the case of a monochromatic radiation is a sinusoidal signal as a function of mirror displacement. In the case of a zero divergence angle, a "pure" sinusoid with a single frequency is produced. However, with a finite divergence angle, the interferogram becomes sinusoidal with a narrow band

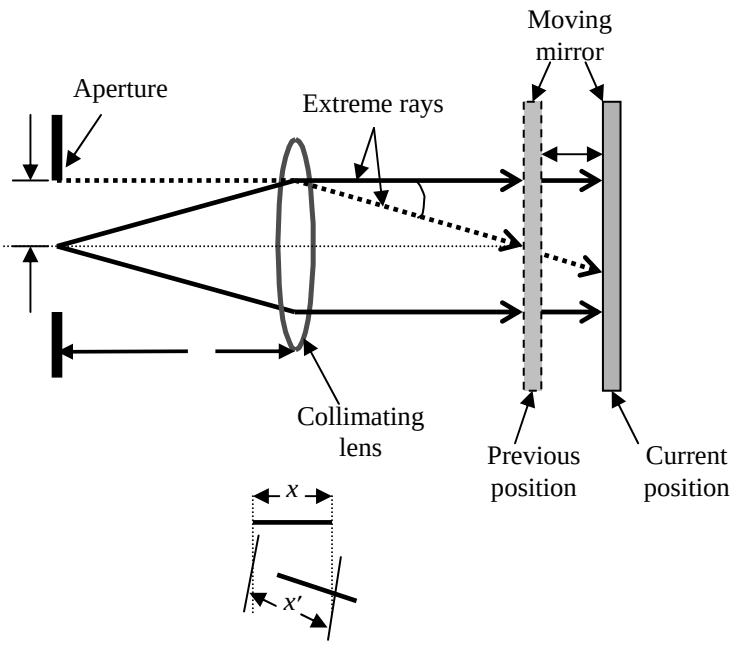


Figure 3.13 A finite divergence angle of the beam entering the interferometer. In most cases involving broadband sources, the beam's divergence angle is limited by the size of an aperture, called the *Jacquinot stop*. The beam splitter and the fixed mirror are omitted in the diagram for simplicity.

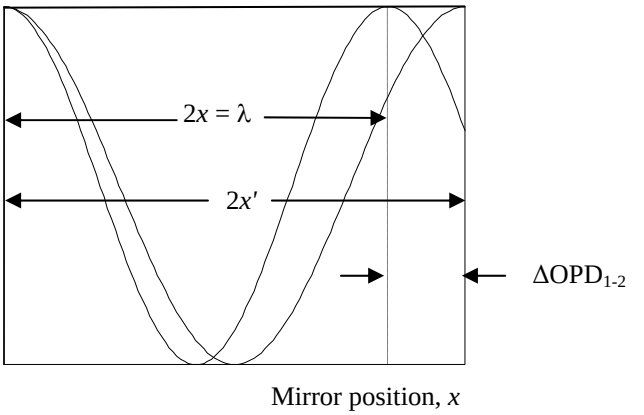


Figure 3.14 Spectral line-width broadening due to beam divergence: two extreme rays undergoing different optical path-lengths.

of frequencies caused by the different path lengths each ray travels. As seen in Fig. 3.14, the two extreme rays undergoing different path lengths result in two sinusoids with different periods. Thus, this monochromatic radiation is analyzed as quasi-monochromatic with a line width defined by the degree of divergence (the extreme rays).

Now calculate the achievable resolving power R as a function of the divergence angle α. From geometry,

$$x' = \frac{x}{\cos(\alpha)}. \tag{3.29}$$

Considering the extreme rays 1 and 2, the difference in the optical path length between them is given by

$$\Delta\mathrm{OPD}_{1-2} = \left|2\left(x' - x\right)\right| = \left|2x\left(\frac{1}{\cos(\alpha)} - 1\right)\right|. \tag{3.30}$$

$$\cos(\alpha) \approx 1 - \frac{\alpha^2}{2}, \tag{3.31}$$

Using the small angle approximation, the difference between the two rays paths can be approximated by

$$\Delta\mathrm{OPD}_{1-2} \approx x\alpha^2. \tag{3.32}$$

Figure 3.14 illustrates the resulting interferogram of the two extreme rays undergoing different optical path lengths. The two rays are out of phase with each other when $\Delta\mathrm{OPD}_{1\text{-}2}$ is equal to $\lambda/2$. Thus, the wavelength λ determines the maximum relative OPD between the two extreme rays that can be "afforded." The shorter the wavelength the more stringent the requirement becomes. For broadband radiation, the shortest wavelength present determines the maximum

value of $\Delta OPD_{1\text{-}2}$, as given by the following equation:

$$\Delta OPD_{1-2} \le \frac{\lambda_{min}}{2} = \frac{1}{2\upsilon_{max}}. \tag{3.33}$$

Combining Eqs. (3.32) and (3.33), the divergence angle is limited by

$$\alpha^2 \le \frac{1}{2x\upsilon_{max}}. \tag{3.34}$$

The spectral resolution, as discussed in the previous section, is also a function of the mirror retardation—that is, $\Delta\upsilon = 1/2x$ (from Eq. 3.27). Thus,

$$\alpha^2 \le \frac{\Delta\upsilon}{\upsilon_{max}} = \frac{1}{R}. \tag{3.35}$$

The divergence angle is determined by the aperture radius. From trigonometry (Fig. 3.13), it is shown that

$$\alpha \approx a/f, \tag{3.36}$$

assuming a small divergence angle, where f is the focal length of the collimating optics. Thus, the size of the aperture needs to be controlled in order to meet the required spectral resolution

$$a \le \frac{f}{\sqrt{R}}. \tag{3.37}$$

Thus, this section closes with an example of an FT spectrometer using a plane-mirror Michelson interferometer for operation in the near infrared with the following requirements:

- Operating spectral range: 12500 – 4000 cm^{-1} (800 – 2500 nm).
- Spectral resolution: 5 cm^{-1}.
- The spectral resolving power: $R = \frac{12500}{5} = 2500$.

  First, the mirror scan distance must be calculated to determine the type of carriage and actuator to drive the moving mirror. From Eq. (3.27),

  $$OPD = \frac{1}{5cm^{-1}} = 0.2cm.$$

  Thus, for a single-sided interferogram, the maximum mirror-scan distance must be 0.1 cm, whereas for a double-sided interferogram, it must be 0.2 cm.

- Next, the collimating optics is designed to ensure that the divergence angle must be kept low enough to meet the resolving power requirement.

Suppose standard 1-in. diameter optics is used, including the plane mirrors. Assuming the use of a collimating lens with an *f*/# equal to 4, the focal length *f* is calculated to be

$$f = 4 \times 2.54\text{cm} = 10.2\text{cm}.$$

Using Eq. (3.37), the maximum aperture radius is then calculated to be

$$a_{\max} = \frac{10\text{cm}}{\sqrt{2500}} = 0.20\text{cm}.$$

- The interferometer throughput using Eq. (1.6) can be calculated as

$$\Theta = \frac{\pi}{R} A_{\text{mirror}} = \frac{\pi}{2500}\text{sr} \times 5.1\text{cm}^2 = 6.4 \times 10^{-3}\text{sr} \cdot \text{cm}^2.$$

3.4 Interferogram Digital Processing

Apodization and zero filling are mathematical techniques generally applied to the interferogram prior to the Fourier-transform routine to improve the output spectral features that are applied. Some of the numerical examples in this book make use of the apodization and the zero-filling techniques to improve the qualitative spectral features. This section serves only as an introduction to these mathematical techniques. Also briefly discussed is another mathematical technique called *phase correction*, which is needed when one uses single-sided interferograms. Further details on these methods can be found elsewhere.

3.4.1 Apodization

As discussed in Sec. 3.3.1, the instrumental line shape for the travel of a finite mirror is that of a *sinc* function. It arises mathematically from the boxcar (or top-hat) function produced by truncating the sinusoidal wave. For some applications, the presence of rather large-sided lobes and their negative values are undesirable.

Apodization in Fourier spectrometry refers to the mathematical processing techniques applied to the interferograms in order to "improve" the instrumental line shape. The name apodization actually derives from its Latin meaning: "cutting off the feet"; the reason will be apparent shortly.

A variety of apodization functions can be used to shape the ILS; the most popular of these include triangular, Gaussian, and exponential. The triangular apodization function results in a $\text{sinc}^2(\upsilon)$ line shape, a Gaussian function in another Gaussian line shape, and an exponential function in a Lorentzian line shape.

How a triangular apodization function (probably the most widely used) affects the ILS is illustrated in Figure 3.15. The processed interferogram is simply the *dot* product of the raw interferogram and the apodization function.

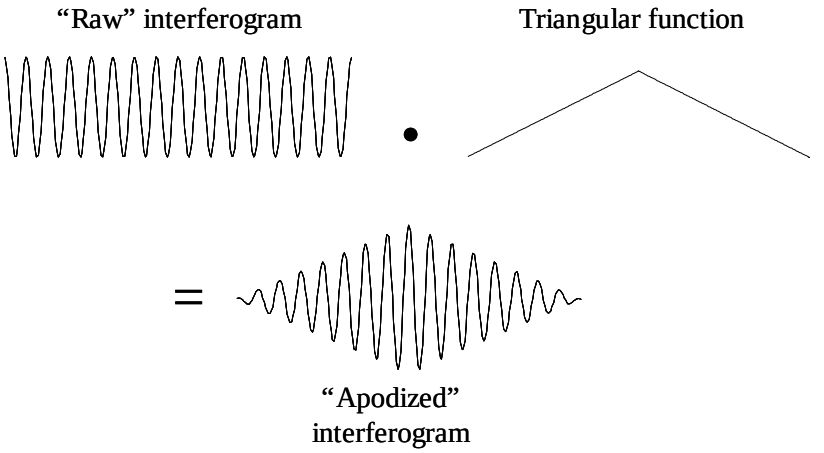


Figure 3.15 Application of the of apodization technique.

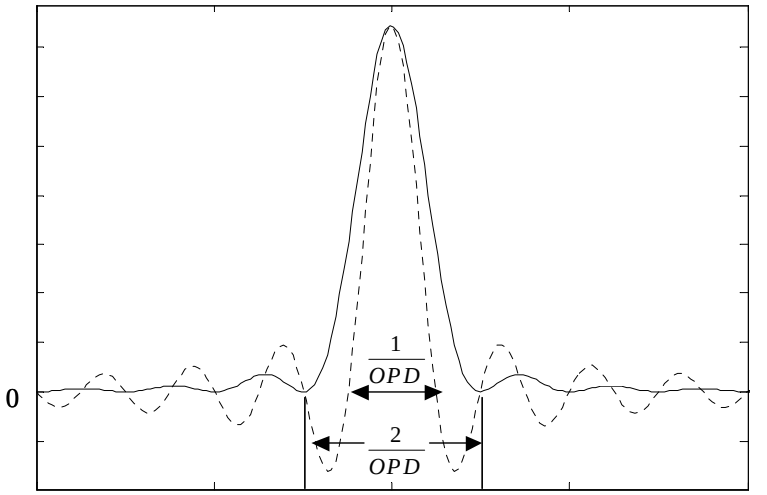


Figure 3.16 Comparison of instrumental line shapes due to a triangular apodization function or a top-hat (no apodization) function. Solid line: triangular function. Dashed line: top-hat function.

The resulting ILS is $\text{sinc}^2(\upsilon)$ with much smaller side lobes and no negative values. The "price paid," however, is in the loss of resolution, as seen in Fig. 3.16. This should be expected because the importance of the data points further away from the zero OPD (by "cutting off the feet") is reduced since those data points are the ones that contain the high-resolution information.

3.4.2 Zero filling

"Zero filling" involves adding zeros (null data) to the tails of the original interferogram. One of the goals is to increase the number of data points until an appropriate number for FFT operation is reached (a power of two). The other goal is to smooth out the resulting spectrum by adding interpolated data points in between the original nonzero-filled spectrum. The outcome of this operation is similar to adding data points in between the original spectral data points through polynomial fitting.

3.4.3 Phase correction

An ideal interferogram is symmetrical about the zero OPD location. Interferogram phase errors may arise due to various reasons. Among them are the misplacement of the centerburst (the interferogram is not sampled at exactly the zero OPD), electronic delays, and the wavelength dependence of the beamsplitter's refractive index. For a double-sided interferogram, these errors do not affect the output spectrum as long as the centerburst is near the center of the recorded interferogram.[5] However, phase corrections need to be performed on single-sided interferograms in order to obtain the correct spectra. Collecting a single-sided interferogram may be necessary in some cases such as when the highest resolution is required out of the possible distance of the mirror travel.

The reader is referred to Bell[5] and Bertie[6] for the details of the phase correction techniques. In practice, the single-sided interferogram includes a short "negative" OPD and the full "positive" OPD to provide the desired resolution. An interferogram signal from the negative OPD is needed to yield a low-resolution-phase spectrum, which is used to correct the spectrum that results from the interferogram from the positive OPD.

References

1. B. E. A. Saleh, M. C. Teich, *Fundamentals of Photonics*, pp. 46, John Wiley & Sons, Inc., New York (1991).
2. J. W. Cooley, J. W. Tukey, "An algorithm for the machine computation of the complex Fourier series," *Mathematics of Computation*, **19**, pp. 297 (1965).
3. D. W. Ball, *The Basics of Spectroscopy*, pp. 105–108, SPIE Press, Bellingham, WA (2001).
4. R. J. Bell, *Introductory Fourier Transform Spectrometry*, pp. 63–68, Academic Press, New York (1972).
5. R. J. Bell, *Introductory Fourier Transform Spectrometry*, pp. 153–168, Academic Press, New York (1972).

6. J. E. Bertie, "Apodization and phase correction," *Analytical Applications of FT-IR to Molecular and Biological Systems*, pp. 25–50, D. Reidel Publishing Company, Dordtecht, Holland (1980).

Chapter 4
Interferometer Alignment Errors

The motion system of the interferometer is probably the most important (and most interesting) subsystem of a Fourier-transform spectrometer. It is essential to understand how errors generated in the interferometer's motion system affect the output spectra. It is the goal of this chapter to portray the links between them.

In any real motion system, "erroneous" motion is unavoidable. Its characteristics and magnitude depend largely on the quality of its mechanical drive components and the design of the overall motion system. Ultimately, the achievable spectral SNR can easily be limited by this subsystem. Figure 4.1 shows three 100% lines from three different mirrors' drive subsystems, having different motion qualities. It can be seen that spectral SNR can be extremely low when the motion quality is poor. Here, motion "quality" refers to the proximity of the path of the actual motion to an ideal one. The ideal motion system in an interferometer has motion in only one degree of freedom, which is in the direction of the OPD.

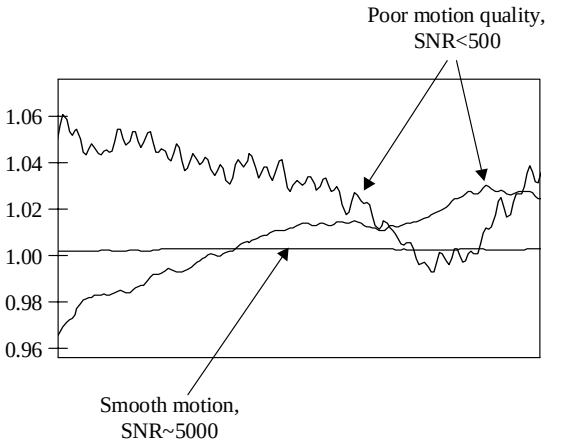


Figure 4.1 Real experimental results showing the effects of moving-mirror-motion quality to spectral SNR. The best 100% line was obtained with an interferometer using a flexural bearing. The other two 100% lines came from an interferometer using poor-quality linear ball bearings, one with a worse feedback-control system than the other.

The first type of spectral error arises from an incorrect angle in the moving mirror and causes optical-path-length errors, also referred to as the interferogram's *phase error* (see Fig. 4.7). This is the result of a mismatch in the phase between the rays that are reflected at the fixed mirror and those reflected at the tilted moving mirror. The second type of spectral error occurs in the interferogram data sampling. Because of the angular errors, the sampling clock signals from the reference laser beam may not fall on the correct mirror retardation position.

This chapter's focus is on the interferogram's phase error or modulation error, and data-sampling error is briefly introduced. A detailed analysis on data sampling error is given in Chapter 6.

For flat-mirror interferometer design, angular-motion errors such as yaw and pitch are the most detrimental to spectral SNR. Figure 4.2 depicts the commonly used terms yaw, pitch, and roll with respect to the motion direction. On the other hand, interferometers using retroreflectors are more sensitive to linear errors perpendicular to the direction of the motion, as shown in Fig. 4.3.

There are many other interferometer design configurations, and each has its own sensitive-motion direction. This discussion and analysis is focused on the flat-mirror configuration; however, the results can be applied to other

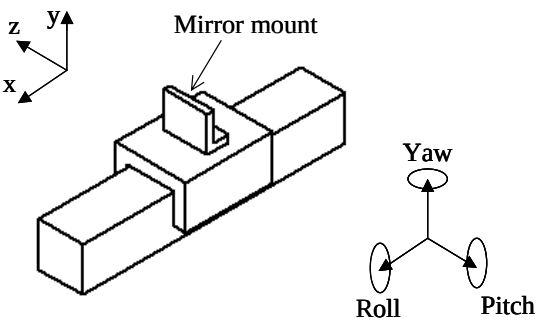


Figure 4.2 Definition of yaw, pitch, and roll for flat-mirror interferometer design. The mirror travels along the *x*-axis.

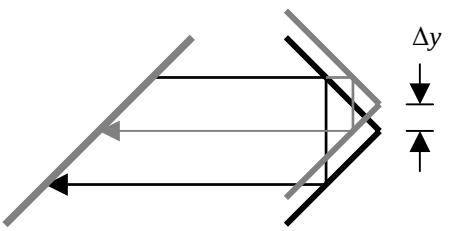


Figure 4.3 Optical-path-length error introduced by an error in motion straightness in interferometers using retroreflectors. The path-length error introduced is equal to $2\Delta y$.

interferometer designs upon simple geometrical conversion. For example, tilts in flat-mirror design result in phase error in the same way as linear lateral motions in an interferometer using retroreflectors. In the former example, optical-path length-error is proportional to the tilt, whereas in the latter it is proportional to the linear displacement.

4.1 Error Characteristics

There are two types of tilt errors: static and dynamic. Static-angular error refers to a constant tilt throughout the mirror's travel and between scans. In this case, the measured spectrum slightly deviates from the "true" spectrum, but is repeatable from measurement to measurement. Thus, the SNR is not affected; the mirror angle can simply be readjusted to provide the maximum modulated signal (indicating optimum alignment). However, serious care should be taken if the interferometer constantly loses alignment over time (hours or days). This usually indicates loose and improper mechanical mountings. Microadjustment screw assemblies often found in microadjustable mirror mounts contain springs that can potentially push the mirror out of alignment if they are not properly locked.

Dynamic angular error refers to tilts with varying magnitudes along the mirror's travel. Here, the nature of the dynamic error can be divided into two categories: repeatable and random. Angular error is said to be repeatable if it is clearly defined as a function of mirror position and does not change over time; that is, the error is repeatable between scans. The opposite is true for random error. Both types of dynamic errors are illustrated in Fig. 4.4. Obviously, repeatable angular error does not degrade the spectral signal-to-noise ratio, while random error does.

It is difficult if not impossible to accurately quantify the effects of dynamic angular motion to the resulting spectrum or the spectral SNR. It depends on various factors such as the exact characteristics of the error itself, the profile of the beam entering the interferometer, the wavelength range, the data-sampling period, among numerous other factors. These factors affect the final spectrum in an interconnected way.

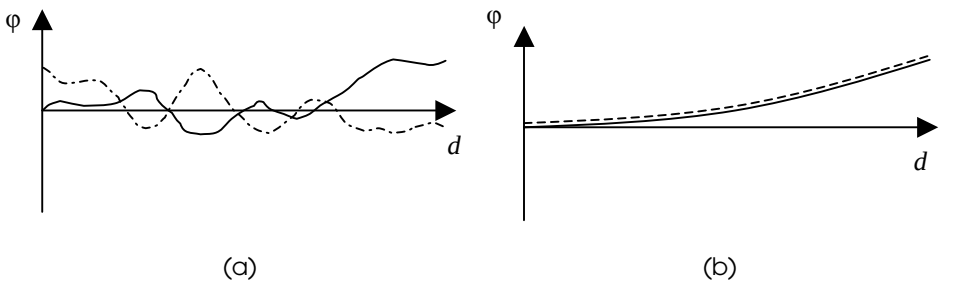


Figure 4.4 (a) Random angular-drive error. (b) Repeatable angular-drive error.

In this chapter, the effects of mirror tilt to the spectral SNR are described in simplified cases with the aim to provide a basic, qualitative understanding to the problem. A basis for performing analyses through numerical simulations is also provided. An example Matlab code is given, which can be expanded to include increasingly complex factors.

Throughout these discussions, keep in mind that pitch and yaw are the two tilt directions to consider. Most of the analysis involves only one for simplicity and clarity. However, it can be easily extended to include the other.

4.2 Interferogram-Modulation Error

For the purpose of this discussion, consider a snapshot in time when the moving mirror has an angular error of φ, as illustrated in Fig. 4.5. The tilt magnitude in the figure is of course exaggerated. Assume that the only effect of the tilt is to alter the OPD, as a function of x (the *sine error*). One might think that this assumption is not valid since it defies the law of reflection (the angle of incidence is equal to the angle of reflectance). However, the rays entering the interferometer have finite beam diameters and they are never perfectly collimated. Thus, a set of rays that fall perpendicular to the tilted moving mirror exists and experiences different values of retardation as a function of x, as illustrated in Fig. 4.6.

The change in retardation for a given tilt angle is

$$\Delta x = y \sin\varphi . \tag{4.1}$$

For a small angle of φ, it becomes

$$\Delta x \approx y\varphi . \tag{4.2}$$

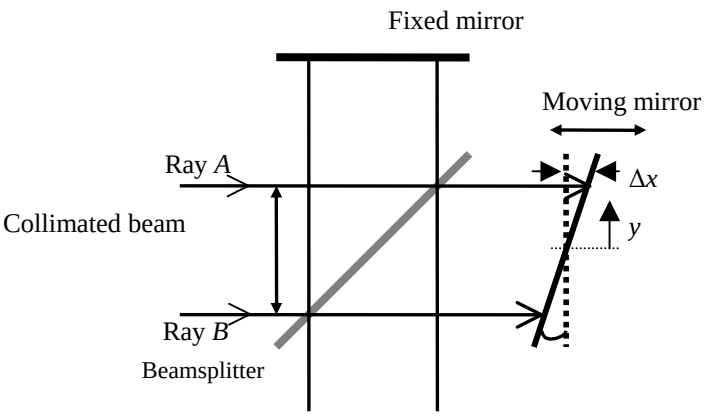


Figure 4.5 Mirror tilt due to an angular error in the mechanical guide.

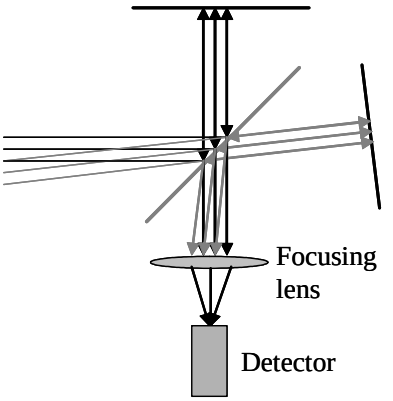


Figure 4.6 Interference of nonparallel rays.

For a wavelength of λ_0, the phase shift between the two extreme rays is (see Eq. 3.15)

$$\Delta\Phi(y) = \frac{2\pi\Delta x}{\lambda_0}. \tag{4.3}$$

The next question is, how does this phase shift affect the resulting spectrum? As described in Sec. 3.2.3 the interferogram of a collimated, monochromatic radiation source is given by

$$I = A\cos(\Phi_1 - \Phi_2), \tag{4.4}$$

where A is the amplitude as determined by the source intensity and the transfer function of the collection optics, and Φ_1 and Φ_2 are the phases of the rays reflected by the fixed mirror and the moving mirror, respectively. Taking into account that the beam is of a finite size (while still omitting the beam dimension in z as defined in Fig. 4.2), the power received by the detector is

$$P = c\int_{-d}^{d} A\cos(\Phi_1 - \Phi_2)dy, \tag{4.5}$$

where c is a proportionality constant, a function of the detector characteristics. Now, due to the tilt, Φ_2 is a function of x. The interferogram function becomes

$$P = c\int_{-d}^{d} A\cos\left[\Phi_1 - \left(\Phi_2 + \frac{4\pi x}{\lambda_0}\right)\right]dx. \tag{4.6}$$

The additional term $2\Delta x/\lambda_0$ represents the additional phase shift in the interferogram introduced by the angular error. Because of the mirror tilt, the interfering rays do not all have the same relative phase with respect to the rays reflected at the fixed mirror, as illustrated in Fig. 4.7. In this figure, the ray in the middle actually has a 90-deg relative phase, and hence results in destructive

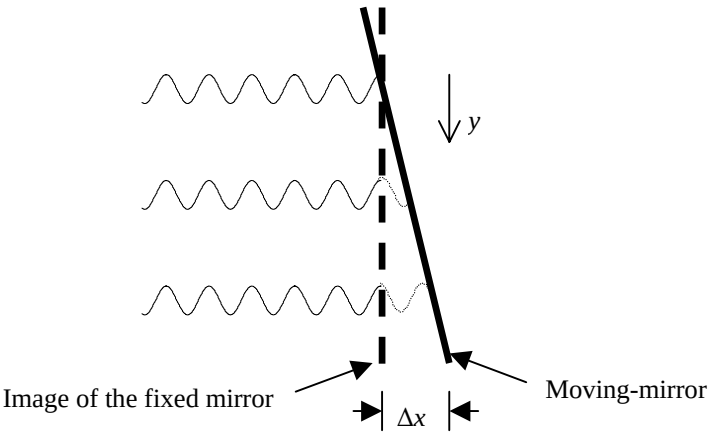


Figure 4.7 Monochromatic interferogram signal affected by phase shift caused by mirror tilt.

interference. Therefore, in general, mirror tilt causes loss of detected radiation due to the phase mismatch. Exactly how it affects the spectrum depends on many factors, as mentioned before, such as the characteristics of the tilt itself. Below are a few examples using numerical simulations to better explain the problem.

Static tilt: In this numerical example, the effect of a static tilt to the output spectrum is analyzed. The following parameters are used in the simulation (see Appendix A for the Matlab program):

λ_0 = 10,000 nm (1,000 cm^{-1})
Tilt: constant at 0.1 mrad
Sampling period: 158.2 nm
Scan distance: 5 mm
Beam diameter: 1 cm
Apodization: triangular
Zero fillings: added to make data points equal to 2^{14}.

As expected, the phase mismatch due to the static mirror tilt causes a power loss in the output spectrum, shown in Fig. 4.8.

The effects of the static tilt to broadband radiation: For a broadband source, the spectrum is not only reduced in magnitude, but its shape is also deformed. This is because the phase mismatch is a function of the wavelength, with the shorter wavelengths affected to a greater degree. This phenomenon is illustrated in Fig. 4.9. This implies that an FT spectrometer covering shorter wavelengths such as the visible or the ultraviolet requires a more demanding mirror-drive quality than a longer-wavelength FT spectrometer. Note also that for a given mirror tilt, the phase mismatch is proportional to the beam diameter. Thus,

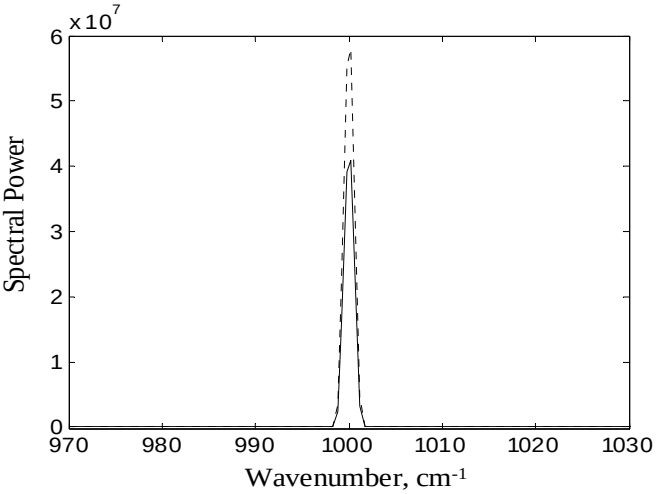


Figure 4.8 Effects of static mirror tilt to the spectrum. Dashed line: spectrum with zero tilt; solid line: spectrum with 0.5 mrad tilt.

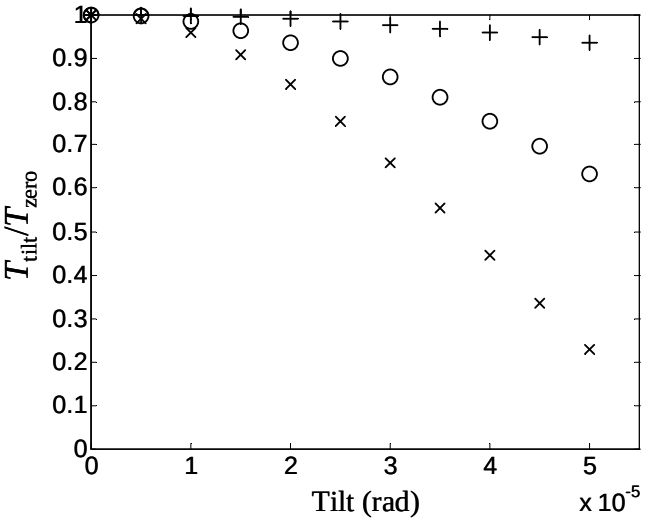


Figure 4.9 Effects of wavelength and tilt angle to transmission loss. + line: υ_0 = 1000 cm^{-1}; o line: υ_0 = 2500 cm^{-1}; x line: υ_0 = 4000 cm^{-1}.

reducing the beam diameter increases the spectral SNR for drive-error-limited instruments.

A common type of dynamic-tilt error is one that increases with travel, mainly caused by error in the linear motion bearing/guide. This type of error is dominant in flexural-bearing stages and air-bearing stages, where random errors due to mechanical contacts are small. A typical figure for air-bearing stages is 2 arcsec *per* 1 cm of travel, which is a result of imperfect straightness of the

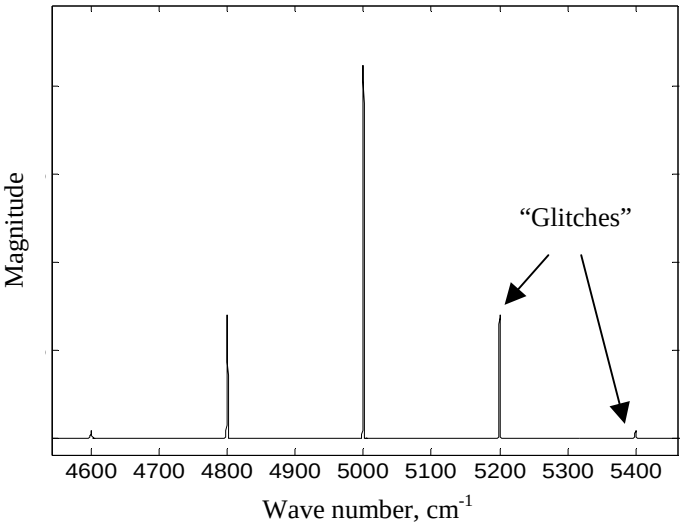


Figure 4.10 Effects of vibrational dynamic tilts of the moving mirror to the transmission spectrum of a monochromatic source.

mechanical guide. Another common type of dynamic error is vibrational, and is due to either external disturbance or by its own amplifier and/or control system.

Numerical simulation is an invaluable tool in complex cases such us these. The Matlab program that simulates the static tilt (given in Appendix A) can be expanded and/or modified to include dynamic angular error.

Dynamic tilt: Another example that demonstrates how dynamic tilts caused by vibration (modeled as sinusoidal) effect the resulting spectrum uses the following parameters.

Parameters: υ_0= 5000 cm^{-1}
Sampling period: 158.2 nm
Scan distance: 2 mm
Beam diameter: 1 cm
Apodization: triangular
Tilt: sinusoidal with 0.02 mrad amplitude at 200 cycles/cm.

The result of this simulation is shown in Fig. 4.10. The vibration of the moving mirror creates secondary peaks or glitches in the transmission spectrum. As with the static case, the effects of dynamic tilt errors are more detrimental to shorter wavelengths. In the simulation, the frequency of the sinusoidal vibration is 200 cycles/cm, resulting in the glitches being shifted 200 cm^{-1} from the spectrum's peak.

Please note that these numerical simulations are based on a few simplifying assumptions. First, the beam profile is assumed to have a top-hat profile, which means that all the rays contribute equally to the total energy detected. Second, the

tilt error only causes phase mismatch. Spectral loss due to the beam missing the detector's active area is neglected.

Despite the simplifying assumptions, this type of simulation analysis is useful if properly used because it illustrates the interdependency of the tilt magnitude, beam diameter, the spectral component, etc. The simulation can also be modified and expanded to include various other factors important to one's specific application.

4.3 Interferogram Sampling Error

Mirror tilts cause data-sampling error if the signal used to "clock" the data acquisition is not perfectly in line with the broadband beam. Detailed analyses of the causes and the effects of sampling error are given in Chapter 6. However, it seems appropriate to introduce the subject here because sampling error may be closely related to the interferometer alignment error, the topic of this chapter.

The most common data-clocking scheme is to use a laser-reference beam. If the reference beam has an offset distance d from the main beam as exemplified by the configuration in Fig. 4.11, a positional sampling error Δd_s is given by

$$\Delta d_s = d \sin\theta , \tag{4.7}$$

for tilt error with an angle θ.

Hirschfeld[1] has suggested that the maximum signal-to-noise ratio achievable with the positional sampling error Δd_s is given by

$$\mathrm{SNR}_s^{\max} = \frac{4}{\upsilon_{\max} \Delta d_s} , \tag{4.8}$$

though he provides no explanation for it.

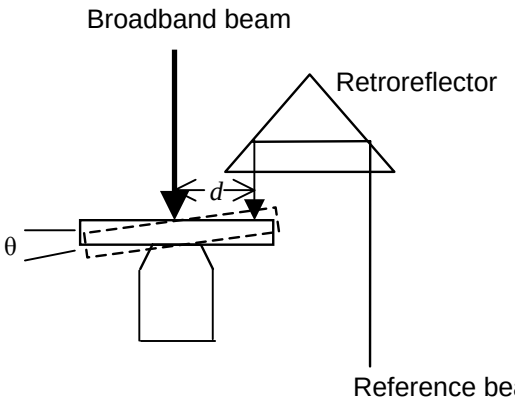


Figure 4.11 The tilt of the moving mirror causes sampling errors due to an offset distance between the laser reference beam and the beam of the main source.

To illustrate the significance of sampling error, consider a spectrometer operating in the near infrared with a maximum wave number of 10,000 cm^{-1} (1,000 nm), an offset distance d of 30 mm, and an angular error of 2 arcsec, which is typical of small air-bearing stages, results in a sampling error of 290 nm. According to Eq. (4.8), the maximum achievable SNR is 14, which is extremely low.

If the interferometer design is inappropriate, the spectral degradation caused by the sampling error may be much greater than the degradation caused by the phase error, as discussed in the previous section. Minimizing or if possible eliminating the offset distance such as the one in Fig. 4.11 is desirable. In reality, many scans are often performed and the interferograms are averaged. This averaging technique tends to improve SNR.

It is important to distinguish once again the difference between dynamic error and static error. If the moving mirror has large but repeatable angular errors, the spectra are repeatable (a high SNR), although they differ from the "true" spectrum.

References

1. T. Hirschfeld, "Quantitative FT-IR: a detailed look at the problems involved," *Fourier Transform Infrared Spectroscopy: Applications to Chemical Systems*, **2**, pp. 193–242, Academic Press, New York (1979).

Chapter 5
Motion Components and Systems

The motion subsystem provides motion to the interferometer's moving mirror, which then creates the interferogram. In this context, the components involved are (1) the actuator, (2) the coupling mechanism (bearing assembly), (3) the position sensor and associated electronics, and (4) the motion controller.

In today's FT spectrometers, the motion systems use feedback control (sometimes called servo control or close-loop feedback control) to regulate the position of the moving mirror. A feedback-controlled interferometric system contains one or more sensors that measure the actual position of the moving mirror at all times, and then performs corrections in real time so that the actual mirror trajectory matches the intended trajectory. A well-designed feedback-control system is insensitive to external disturbances. Figure 5.1 shows a generic architecture of a feedback-control motion system using a digital motion controller.

In this chapter, the types and the design aspects of some of the actuators, bearings, sensors, and motion controllers that are appropriate for use in FT spectrometers are discussed.

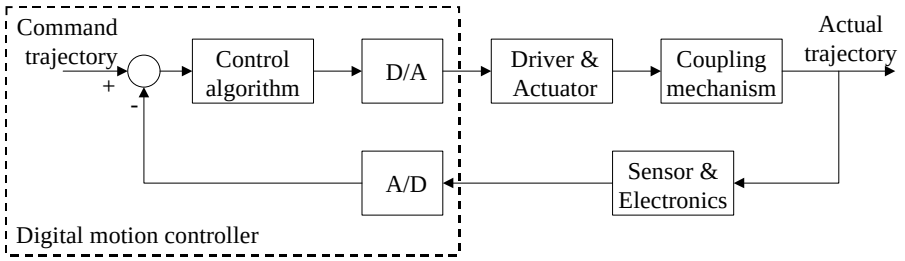


Figure 5.1 A generic architecture of a computer-controlled feedback-control motion system.

5.1 Actuators

5.1.1 Actuator types

The function of the actuator is to convert electrical energy (voltage multiplied by current over time) into mechanical energy (force multiplied by displacement). Two types of actuators commonly used in precision motion applications today are electromagnetic and piezoelectric actuators. An electromagnetic actuator generally uses a permanent magnet and a coil winding that produces a force proportional to the current applied to the coil. A piezoelectric actuator, on the other hand, provides mechanical displacement through material elongation or deformation as an electrical voltage is applied.

5.1.2 Performance specifications for actuators

In this section, some of actuators' performance specifications important to FT spectrometer design are discussed.

Travel range. The first consideration for a linear-motion system should be the travel range. As discussed in Sec. 3.3.1, the maximum travel of the moving mirror directly affects the achievable spectral resolution of the instrument. Single-phase electromagnetic actuators, such as voice-coil actuators (those used in loud speakers) have maximum travel on the order of millimeters. Piezoelectric actuators, on the other hand, can achieve displacement (through strain) only on the order of 0.1% of its length. Thus, a 5-cm long piezo material provides a displacement of approximately 50 μm. This short travel range usually necessitates the use of some sort of mechanical leverage system to amplify the displacement.

Resolution. Referring to the minimum amount of the incremental force produced by the actuator, the resolution is limited by the amount of noise such as the electrical noise and friction or stiction. Multiphase-electromagnetic actuators have a certain noise called a cogging force. This force results from the alignment of the magnet and the coil, and can cause major problems in a continuous-scan interferometer when a relatively slow speed is supposed to be held constant. Of course, what matters in the end is the resolution of the moving mirror's displacement, which is a function of not only the actuator but also the motion system as a whole.

Thermal characteristics. Heat can be a serious problem that limits the accuracy and repeatability of a precision-motion system. Heat causes certain components to expand, which may cause the OPD of the two interfering beams in the interferometer to drift. Furthermore, the wavelength of the laser reference depends on the surrounding gas's index of refraction, which depends on its temperature. To give an idea on the error magnitude due to the change in the medium's refractive index, a 1-part-per-million displacement error is caused by a

1°C change in temperature. An advantage of piezoelectric actuators over their electromagnetic counterparts is the fact that they produce negligible heat.

Maximum force: Given the mass, friction, and stiffness of the drive assembly, the maximum force determines the allowable speed, acceleration, and deceleration that are important for both the step-scan and the continuous-scan operations. The difference between these operations is explained in Sec. 6.1. For a step-scan operation, one would try to minimize the *step-and-settle* time (see Fig. 5.2), whereas for a continuous-scan operation, one would try to minimize the "turn-around" time (see Fig. 5.3).

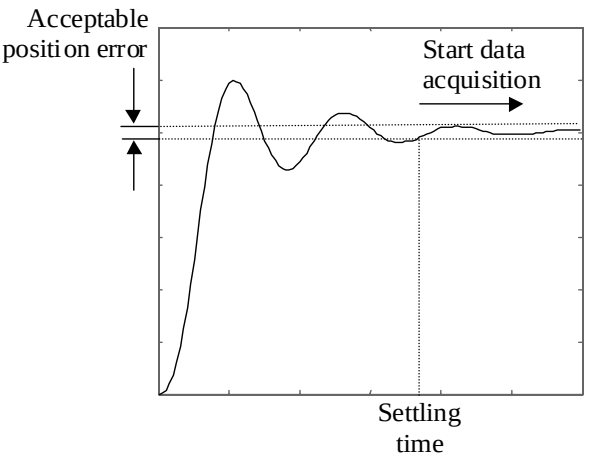


Figure 5.2 A trajectory of the moving mirror under a step-and-scan operation. The "settling time" is the waiting period before data acquisition can start so that the sampling error is within an acceptable value.

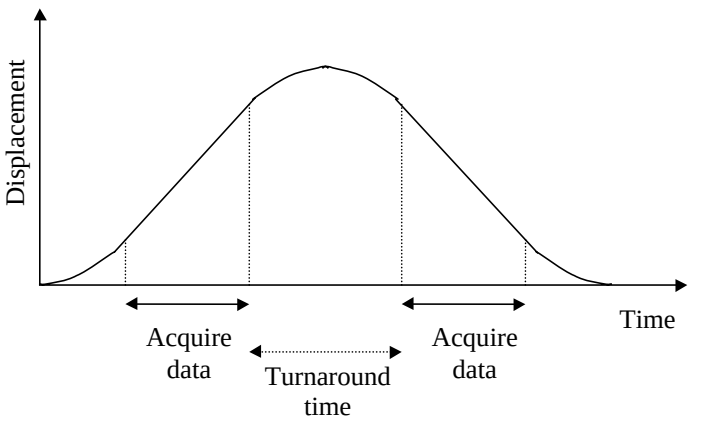


Figure 5.3 The moving-mirror displacement under a continuous-scan operation. The "turn-around" time is an "unproductive" period that is used to decelerate and accelerate the moving mirror. Data acquisition can take place only when the velocity is constant.

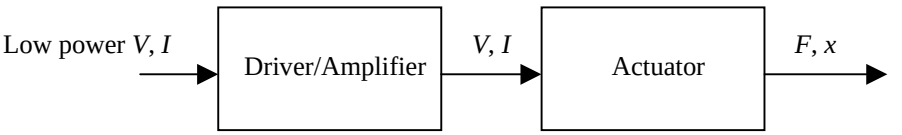


Figure 5.4 Block diagram driver-actuator system.

Bandwidth. An indirect measurement of how fast the actuator can react and follow the commanded trajectory, bandwidth is given in units of Hertz. Although motions in an FT spectrometer are relatively slow (< 100 Hz), a large bandwidth is desirable to be able to compensate for disturbances. Since the motion system is under servo-control (or it should be), it tries to correct any positional error that arises. However, it is not able to do so efficiently if the actuator cannot respond fast enough.

5.2 Driver and Amplifier

The main function of the driver, or the amplifier unit, is to convert the low-power electrical signal from the motion controller to a higher power signal to drive the actuator. In the case of electromagnetic actuators, the amplifier unit produces current (I) proportional to the voltage (V) command. The electromagnetic actuator then produces force (F), which is proportional to the input current. Figure 5.4 illustrates the process.

Of the actuator's performance factors mentioned in the previous section, resolution, maximum force, and bandwidth are partly dependent on the amplifier. The noise in the electric-signal output affects the actuator's resolution, especially if it is within the actuator's bandwidth. For example, if the actuator's bandwidth is 1 kHz, noise with a frequency of 100 kHz affects the overall system much less than noise with a frequency of 10 Hz. To ensure optimum performance, the amplifier's maximum power and bandwidth should be chosen so that it exceeds that of the actuator's.

5.3 Bearings and Coupling Mechanism

The main function of the bearing mechanism in interferometer design is to provide "smooth" motion in one degree of freedom while preventing motion in the other five degrees of freedom. Various types of bearings can be employed in FT spectrometers, for both linear and rotary motion. Here, the focus is on the linear type, the most commonly employed for FT spectrometers.

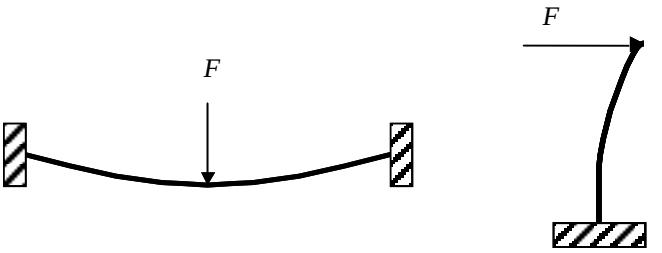


Figure 5.5 The principle of operation for flexural bearings. Displacement in the direction of motion is provided through the deflection of the bearing material.

5.3.1 Bearing types

Bearings can be classified into two categories.

(1) *Bearings that rely on mechanical contact between the moving elements*. They include sliding-contact (friction) bearings, ball bearings, and flexural bearings. The oldest and the simplest type of bearing is the sliding-contact type, where the base and the plate move relative to each other with the help of lubricants. Ball-bearing stages use small, precision-machined steel balls to provide relative motion between the base and the plate. Flexural stages rely on the elasticity of a material to provide displacement under force, as illustrated in Fig. 5.5. Figure 5.7 (page 54) shows a commonly employed flexural bearing design for motion systems that require one degree of freedom.

(2) *Bearings that do not rely on mechanical contact between the moving elements*. They include aerostatic bearings (air bearings), hydrostatic bearings, and magnetic bearings. Air and hydrostatic bearings use a thin film of pressurized air and liquid (usually oil), respectively, to "float" the moving plate above the base, therefore reducing friction drastically. Thus, they require a constant source of pressurized fluid. Magnetic bearings, on the other hand, use electromagnetic force to float the moving plate. Real-time closed-loop control of the injected force is needed to provide stable positioning of the plate. See Ref. [1] for excellent reviews on these various types of bearings.

5.3.2 Performance specifications for bearings

In this section, some of the bearing's specifications important to FT spectrometer design are discussed.

Range of motion. Although the bearing mechanism itself does not always limit the maximum travel, the linear mechanical-motion assembly, which is often called the *stage* or the mechanical stage, does have travel limitations. Many off-the-shelf ball-bearing and air-bearing stages provide > 20 cm distance of travel (thus they are able to provide < 0.05 cm^{-1} resolution). In contrast, flexural bearings are usually limited to a few millimeters of travel. As a rule of thumb,

travel of a flexural-bearing stage is generally limited to one tenth of the bearing size.[1] For some low-resolution applications, such as spectroscopy in the near infrared, this type of bearing may be superior to others because of their smooth repeatable motion as discussed in the following paragraphs.

Angular errors. It is the function of the mechanical stage to maintain the moving mirror's angle perpendicular to the fixed mirror throughout its travel. Using flat mirrors in the Michelson configuration, the sensitive angular errors are those of pitch and yaw. Roll motion about the center of the mirror (and thus the beam) does not affect the OPD. A stage using a contact bearing generally has larger and more unpredictable angular errors due to manufacturing imperfections and wear of the contacting elements. High-precision ball-bearing stages with a few centimeters' travel have pitch and yaw motions in tens of arc seconds, compared with less than a couple of arc seconds for air-bearing stages. Their error characteristics are also more unpredictable, resulting in complex dynamic alignment errors. A properly designed flexural bearing may compete with the more expensive air-bearing stages in terms of the angular error performance. Both air-bearing and flexural-bearing stages generally have repeatable angular error characteristics.

Displacement resolution. Measuring the ability to follow a small increment of motion or displacement is important for both the step-scan and the continuous-scan operations. To the former, the achievable motion resolution determines the minimum sampling period, which in turn determines the highest wave number (shortest wavelength) the spectrometer can synthesize, as is discussed in Chapter 6. To understand the significance of the bearing resolution in the continuous-scan operation, it is important to recognize that bearing resolution is normally limited by nonlinear and irregular friction forces that cause stick-slip occurrences. Such occurrences make constant-velocity motion difficult to achieve. Velocity variations during data sampling in a continuous-scan-operated interferometer may result in sampling error (discussed in Chapter 6). Surface and finish imperfections of the bearing's contacting elements are largely responsible for generating the irregular friction forces during motions. Therefore, noncontact bearings are capable of smoother motion, and thus have better resolution than the contact bearings. Flexural bearings provide an exception. Although they are considered to be contact bearings, the actual contact elements are the millions of atomic bonds that stretch and contract during the elastic motion. Due to the averaging effect, flexural bearings can achieve atomically smooth motion.

Stiffness. High stiffness in the other five degrees of freedom is desirable, as it would make the stage insensitive to disturbances. Stiffness is given in units of newtons per meter for the linear directions and in units of newton-meter per radian for the angular directions. For mechanical contact bearings, however, an increase in stiffness usually results in poorer resolution due to the increase in the generally unpredictable static and low-velocity friction forces. For example,

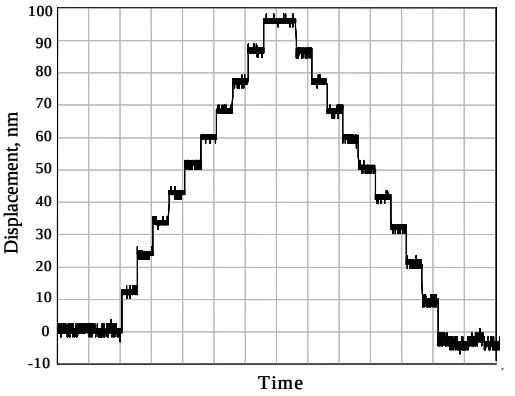


Figure 5.6 Step-scan operation of a well-designed air-bearing stage, with a commanded step size of 10 nm. Position sensor: a linear optical encoder with a resolution of 31 picometer (after interpolation). The target was measured by a heterodyne-laser interferometer with a resolution of 1.25 mm.

stiffness in a linear ball-bearing stage depends on the magnitude of the *preload*[1] (adjustable by the user or set by the manufacturer). Higher preload increases the stiffness, while it decreases the resolution because of the higher friction between the rolling elements.

5.3.3 Discussion on bearings

Most of the "high-end" FT spectrometers employ air-bearing stages. High-end instruments generally refer to those that provide high-spectral resolution, fast scanning, and high-spectral SNR. Air-bearing stages provide a superb combination of smoothness, small errors, and a long travel range. Figure 5.6 shows a step-scan operation of an air-bearing stage with a 270-mm travel range, and a step size of 10 nm. Such a combination of long travel range and superb resolution capabilities would be impossible to achieve with contact-bearing stages.

The drawbacks of air-bearing stages include their relative expense and the requirement of a constant source of pressurized, filtered air. Magnetic-bearing stages can potentially offer a similar performance to the air-bearing stages. However, they are not as popular because they cost more and are more complex owing to the active stabilization control of the floating plate.

The inherent simplicity of flexural stage design and its ability to provide smooth motion should be great enough motivation to explore this device for some FT spectroscopy applications. The flexural stage shown in Fig. 5.7 can be

[1] Preload is the force applied to a bearing assembly to remove the axial play or movement and to further compress the bearing to provide both axial and radial stiffness. The direction of the applied force is orthogonal to the motion axis.

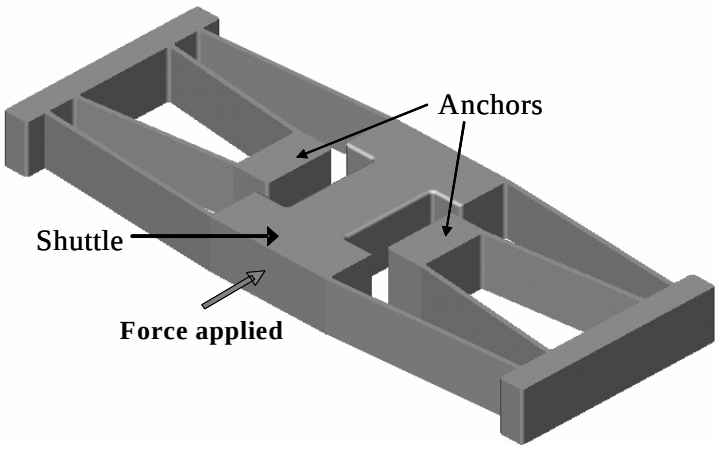


Figure 5.7 A solid model of a modified *folded-beam* suspension design, one of the commonly employed flexural-bearing configurations for a one-degree-of-freedom motion system.

manufactured from a single steel plate, cut with either a water-jet machining method or a wire electrical-discharge-machining (EDM) method. The design shown in the figure is a modified *folded-beam suspension*, called a "double V-beam suspension."[2] They can certainly be excellent choices for spectrometers designed for low-resolution measurements. The main advantages of flexural bearings include low cost, negligible friction, and the need for little or no maintenance.

The simple deflection-based design of flexural bearings lends itself to a flexibility that no other bearing types offer. Various flexural stages with "clever" mechanical designs that minimize errors and amplify displacements can be laid out and easily manufactured. For example, in order to minimize angular error [like the one shown in Fig. 4.4(b)], the force can simply be applied at a different position, maybe offset from the center as illustrated in Fig. 5.7. In the MB series FTS made by ABB Bomem, Inc. of Quebec, Canada, a flexural bearing's limited motion (its largest single drawback) is compensated for by the use of rotary-to-linear motion amplification, shown in Fig. 5.8. The resulting maximum linear OPD is 2 cm (1-cm optical path difference before and after the centerburst), resulting in 0.5-cm^{-1} spectral resolution according to Eq. (3.27).

There are commercial-off-the-shelf linear-motion devices incorporating flexural stages called *shakers* and are used for vibration testing. Some examples are the Mini-shaker type 4810 (Bruel-Kjaer, Denmark) and the Vibrator V200 series (Ling Dynamic Systems, CT, USA). They use a permanent field-magnet voice-coil actuator, with the moving coil suspended by radial flexure springs, providing up to 6 mm of travel.

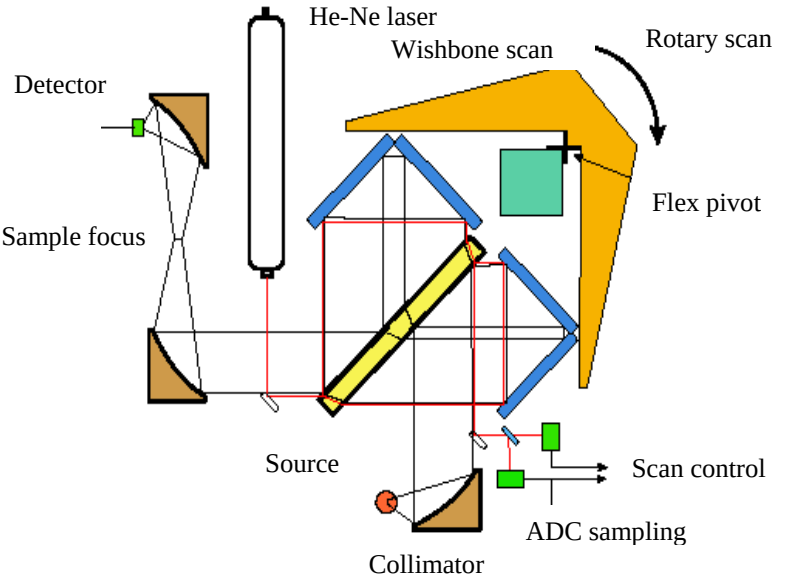


Figure 5.8 Simplified optical layout of Bomem MB series FT spectrometer, highlighting its unique interferometer design. It uses a flexural bearing that provides a rotary motion. (Reprinted by permission from ABB Bomem, Inc., Quebec, Canada).

5.4 Position Sensor

The position sensor in the FT spectrometer's interferometer usually provides a position-feedback signal to the motion-control board and a sampling clock signal for data acquisition for the interferogram. There are two types of position sensors: *analog*, where the output (usually in voltage) is proportional to the distance measured; and *incremental*, where the output is a series of pulses; the number of pulses is proportional to the distance traveled.

Examples of analog-type position sensors include capacitance sensors, linear variable differential transformers (LVDTs), Hall-effect sensors, and optical confocal sensors. Incremental sensors include linear (and rotary) optical encoders and laser-interferometric sensors.

In a continuous-scan interferometer, the second function of the position sensor is to provide a sampling clock signal, and generally requires repetitive digital pulses along the mirror displacement. The digital pulses are used by the data-acquisition system to time or to trigger the analog-to-digital conversions in real-time, with minimum delay. Thus, for a continuous-scan operation, using an incremental-type sensor makes more sense. Otherwise, two separate sensors are needed: one for motion control and another for the data-acquisition clock signal.

For a step-scan interferometer operation, on the other hand, one does not need a real-time data-acquisition clock signal. The analog-to-digital data conversion and collection can be programmed to start at a certain time after the position command signal is generated or after the moving mirror is settled within a certain displacement error window (see Fig. 5.2).

5.4.1 Laser interferometer

Most commercial FT spectrometers today use HeNe-laser interferometers as position sensors. There are two types of laser interferometers: *homodyne* and *heterodyne*. A homodyne interferometer uses a single-frequency laser source, whereas a heterodyne interferometer uses a laser source with two close frequencies.

The homodyne interferometer using a HeNe laser as its source is a commonly used position sensor in FT spectrometers. It is known that with a simple Michelson setup, the moving mirror's relative displacement can be determined by counting the sinusoidal pulses. The speed information can thus be obtained by measuring its frequency. However, in order to determine the direction of the travel (hence the position and the velocity), additional optics is needed. A typical optical layout to accomplish such tasks is shown in Fig. 5.9.

A linearly polarized laser beam passes through a $\lambda/4$ retarder (a quarter-wave plate) and it becomes circularly polarized. In order for this to occur, the incoming beam's polarization direction needs to be at 45 deg relative to the retarder's principal axis. Circularly polarized light has orthogonal fields that are 90 deg

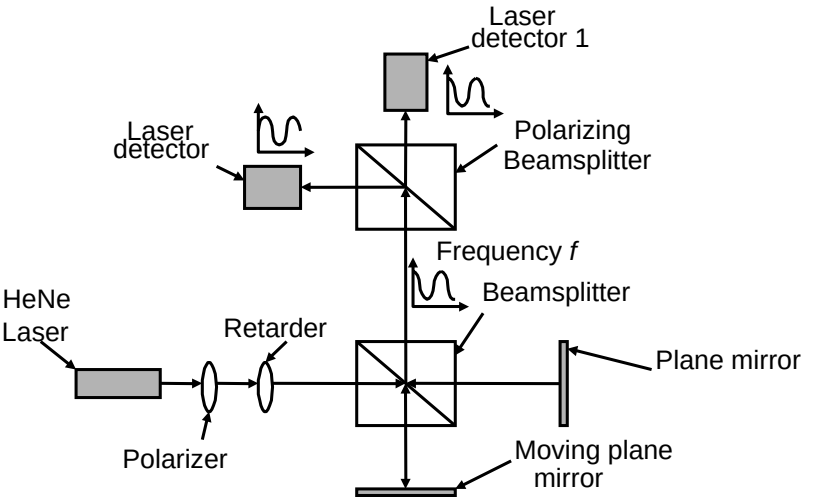


Figure 5.9 Homodyne-laser interferometer with quadrature detection.

apart in phase, for example, the *x*-component leading the *y*-component by λ/4. A polarizing beamsplitter then splits the output beam into two orthogonal fields, each sensed by a detector. As a result, depending on the direction of the mirror's travel, detector 1 leads or lags detector 2. This phase information can then be used to deduce direction and therefore the position of the mirror. Most commercial motion-control boards contain two channel-encoder inputs that accept digital quadrature-encoder signals and automatically perform directional measurements or up and down pulse counting.

A heterodyne interferometer measures mirror displacement by measuring the phase change due to the Doppler effect, similar to the way police radar measures the speed of a car. In order to do this, however, the radiation's temporal frequency needs to be made much smaller (remember that light's temporal frequency is much too high to be detected with today's detectors and electronics). This is achieved by using a source that contains two frequency components located close to each other, creating a *beat* phenomenon whose amplitude variation can be made much slower than the radiation frequency itself (in the order of megahertz), as illustrated in Fig. 5.10. The envelope frequency is given by $f_1 - f_2$, the difference of the two frequency components of the source radiation.

Figure 5.11 shows a simple schematic of a heterodyne interferometer configuration. The two frequency components have linear polarization, orthogonal to each other, which are then separated at the polarizing beamsplitter. The recombined radiation has a beat frequency of

$$f_{\text{beat}} = f_1 - (f_2 + \Delta f), \tag{5.1}$$

where the Δf term is due to the Doppler shift. Its polarity is dependent on the direction of the motion. This signal is then electronically "compared" with the reference signal $f_1 - f_2$ (the beat frequency corresponds to zero velocity). Readers

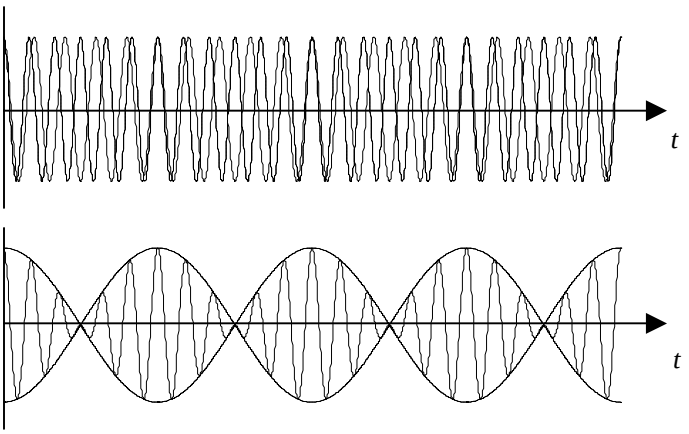


Figure 5.10 Beat phenomenon caused by the superposition of two monochromatic waves having two slightly different frequency components.

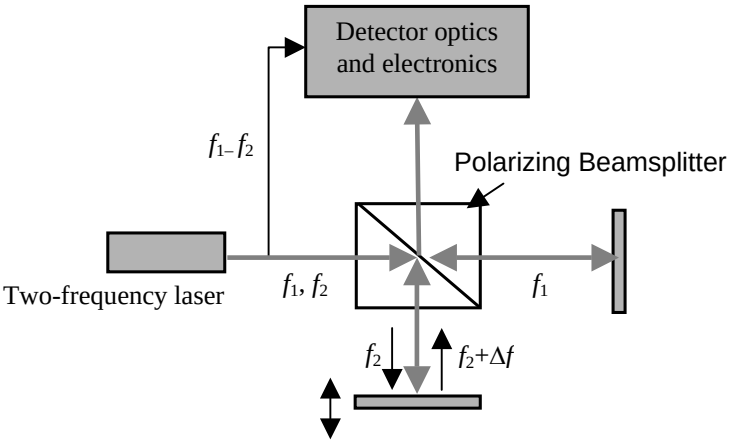


Figure 5.11 Simplified schematic of a heterodyne interferometer.

are referred to Ref. [1] for complete mathematical treatment of the Doppler shift effects. A phase detector is then used to measure the phase between the reference and the measured signals.

A few aspects of the heterodyne interferometer make it superior to the homodyne interferometer. Since the displacement information is carried on an ac signal rather than a dc signal, it is less sensitive to laser power fluctuations, ambient light, and various other noise affecting dc measurements. In addition, only a single detector is needed to measure both the displacement and the direction, thus simplifying alignment. However, it requires a highly stabilized two-frequency-laser source and more intricate detector electronics.

5.4.2 Sensor specifications

Resolution: Assuming that the sensor signals are used to clock the interferogram data acquisition, the resolution of the sensor determines the minimum sampling period and, in turn, the maximum wave number observable, as discussed in Sec. 6.2.1. The resolution of the position sensor also affects the minimum incremental step of the moving mirror (important for the step-scan operation) and the variations of the velocity (important for continuous-scan operation). A well-designed feedback-control motion system can have positional "jitters" as low as ± 1 encoder count (within ± 1 resolution unit). With appropriate electronics, an interferometer using a HeNe laser can achieve resolution on the order of a few nanometers. Some heterodyne laser interferometers can even achieve subnanometer resolution.

Repeatability: Irreproducibility of the position sensor leads directly to interferogram-sampling error. Repeatability of a laser interferometer depends on its electronics, the stability of the laser source, as well as the environment of the beam path. These factors are analyzed in more detail in Sec. 5.5.

Accuracy: Sensor accuracy affects the output spectrum since the wavelength information concerning the source radiation is obtained directly from the OPD reported by the sensor. The relationship between the accuracy of the position sensor and the accuracy of the output spectrum is discussed in Sec. 6.3.1.

5.5 Homodyne Interferometer Sources of Error

In this section, the potential sources of errors in the laser-interferometer sensor are analyzed and discussed. Throughout the analysis, only the repeatability is considered, not the accuracy. Thus, the relevant question is, how much does the sensor vary after it is calibrated?

As long as the error is linear, accuracy can be made irrelevant in most cases. To clarify this point, consider a case where the sensor gives a positional reading that deviates from the true value by a factor of κ, as illustrated in the following equation:

$$x_{\text{read}} = x_{\text{true}} \times \kappa . \tag{5.2}$$

As discussed before, this readout signal is used for motion control and as a data-sampling clock. Since both of the processes are done in software, the error could easily be accounted and compensated for. In practice, spectrometers are calibrated by measuring known spectral peaks. For grating-based instruments, calibration concerns the angular accuracy of the rotary stage, and for FT spectrometers, calibration concerns the linear positional accuracy of the moving mirror.

5.5.1 Source wavelength variations

The laser source's wavelength stability mainly depends on the stability of the cavity length. Free-running or unstabilized HeNe lasers are stable to within one part in 10^5 (approximately within 0.006 nm). Commercially available frequency-stabilized HeNe lasers have wavelength stability on the order of one part in 10^8–10^{10}. For example, the Melles Griot catalog[3] claims ±1 MHz frequency stability at 473.61254 THz nominal frequency. (Where c = 2.997925e8 m/s, the nominal wavelength is 632.99105 nm in a vacuum or 632.81644 nm in air with a refractive index of 1.00027593). Thus, the stability in wavelength units is ±1.3e – 6 nm.

5.5.2 Refractive index variations

The wavelength of light is affected by the refractive index of the medium in which the light waves travel. Environmental factors that affect the refractive index of the air are temperature, humidity, pressure, and molecular composition. The relation between these factors is given empirically by Schellekens[4] as

$$n = 1 + \frac{2.879294 \times 10^{-9}\left[1 + 0.54 \times 10^{-6}\left(C - 300\right)\right]P}{1 + 0.003671 \times T} - 0.42063 \times 10^{-9} \times F \qquad (5.3)$$

where C is the CO_2 content in parts per million (ppm), F is the water vapor pressure in Pa, P is the air pressure in Pa, and T is the air temperature in degrees Celsius. It is found that the refractive index is most sensitive to temperature change, which is on the order of 1×10^{-6} /°C.[1]

5.5.3 Laser detector errors

Laser detector errors include amplitude-fluctuation noise and phase delays from the detector element itself as well as its amplifier circuit. Their contribution to the position error is highly system dependent, including the displacement resolution and the required speed of the moving mirror. For example, in a low-resolution interferometric system that relies only on the zero crossings of the reference-laser-modulation signal (see Fig. 5.12), amplitude fluctuations would not cause significant erroneous displacement readings. These systems include FT spectrometers for long wavelengths, such as measurements in the far infrared. The relationship between the required resolution of the position sensor and the shortest measurable wavelength is discussed in Sec. 6.2.

5.5.4 Polarization errors

A homodyne interferometer relies on the phase difference between the two orthogonal linearly polarized components to determine the position of the , as shown in Fig. 5.9, to create a 90-deg phase delay on one of the orthogonal components. Deviation from this value results in periodic, erroneous position information, as illustrated in Fig. 5.12.

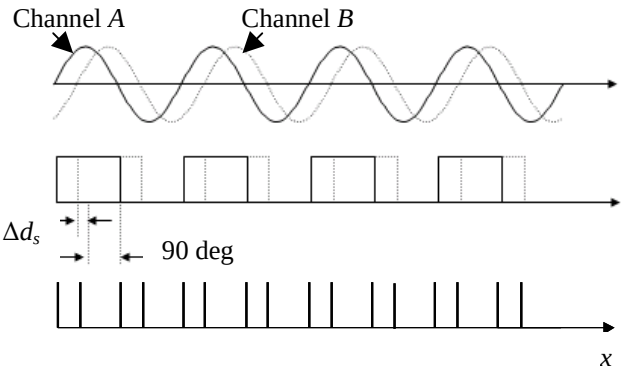


Figure 5.12 Periodic position error caused by imperfect polarization. Perfectly circular polarization results in 90-deg phase between the two channels.

5.5.5 Alignment errors

There are two main types of alignment error: *cosine* and *sine* (Abbe). The latter occurs when the position information is not obtained directly at the "tool point," which, in an FT spectrometer, is generally located at the center of the mirror. Sine error has been discussed throughout this book, for example, in Sec. 4.3 where sampling error was discussed. However, cosine error occurs when the measuring laser beam is not parallel to the axis of motion. For a deviation angle α, the displacement error $(x' - x)$ is equal to $x[1/(\cos\alpha - 1)]$, where x is the displacement of the moving mirror (see Fig. 5.13).

5.6 Actuator, Sensor, and Mirror Mounting

Using motion components with high-accuracy, high-resolution, and low-noise characteristics do not result in a well-built system if the mechanical assembly incorporating them is not well designed. Frequently, large positional feedback and sampling errors arise from "improper" mechanical design, even though high-quality components are used. The discussion in this section includes positional errors that may occur when the actuator, the sensor, the mirror, and the drive system's center of mass are not optimally placed.

5.6.1 Actuator location vs. center of mass

An ideal configuration is one where the point of actuation is inline with the moving system's center of mass. In this case, no momentum is created as force is applied. All types of mechanical stages have finite stiffness, which at the submicrometer level may act like rubber. Angular errors increase with the offset distance and magnitude of accelerations of the moving mass, and decrease with a stiffer bearing. A detailed, dynamic analysis is beyond the scope of this book; however, the moral of this discussion is that one should design the interferometer in such a way that the actuator is "pushing" at a location as close to the center of mass of the moving system as possible.

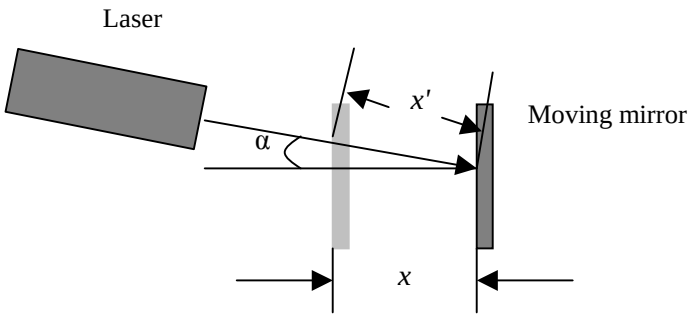


Figure 5.13 Illustration of cosine error caused by misalignment of the laser source.

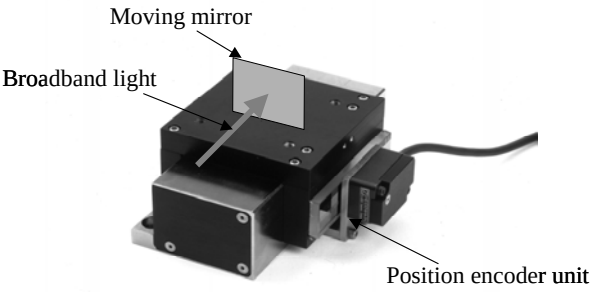


Figure 5.14 Positioning device with an optical encoder as its position sensor. The offset between the mirror and the sensor location may produce significant positional and sampling errors.

5.6.2 Sensor location vs. mirror location

The only positional information of interest is that of the mirror—more specifically, where the center of the broadband light beam is incident. Thus, the measurement point should be as close to it as possible, or preferably, right on it. This is one of the advantages of using a laser interferometer as a position sensor, as it can be designed to directly measure the point of interest. Other sensors are generally not as versatile in their mounting configurations due to their physical size. Figure 5.14 shows a mirror drive system using an optical encoder where there is an offset distance between the mirror and the position sensor.

5.6.3 Sensor location vs. actuator location

The position of the actuator relative to the measured location is important for motion control. Improper placement of the sensor with respect to the actuator results in low-bandwidth-motion systems that are sensitive to external disturbances. An example of such a system is shown in Fig. 5.15. As seen in the figure, there is a rather large offset distance between the sensing location (the laser-interferometer signal is reflected at the smaller mirror) and the voice-coil actuator, making it difficult to achieve “tight” motion control as an oscillatory or even unstable behavior results. This system may be sufficient for long-wavelength applications (far infrared) where fine-positioning resolution is not essential.

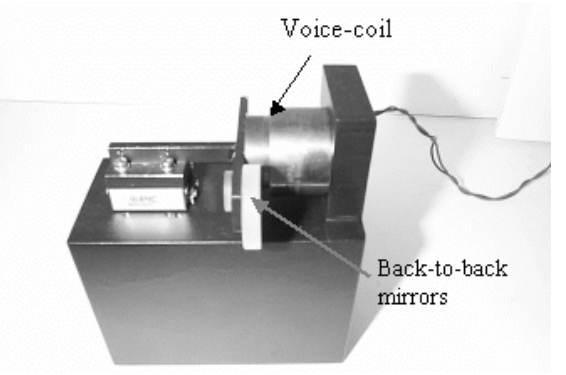


Figure 5.15 Picture of a positioning device using a ball-bearing stage and a voice-coil actuator. The laser interferometer beam is reflected at the smaller mirror. The sensing position is not located in line with the actuator, resulting in low bandwidth.

5.7 Digital Motion Controller

In practically all modern systems, the controller is a digital computer. There are many advantages of using digital control compared to using the analog equivalent; the most important of which is its greater flexibility. Changes in the control parameters as well as in the algorithms to meet the desired response can be performed by the software rather than by the hardware. This can result in minutes rather than hours of designing and optimizing the system. In addition, many tasks, loops, and motion axes can be controlled and/or monitored by a single digital controller through time-sharing. For example, a digital controller can also monitor and make adjustments for background air temperature, pressure, light source intensity, temperature of the detector elements, etc.

Today, there are numerous off-the-shelf controller boards, and making a selection can be confounding. At the minimum, the motion-control board should contain an encoder input, an on-board microprocessor, and a D/A output channel. This is generally enough to control one axis. If an analog position sensor is used, a control board that contains an A/D channel is required to convert the analog position information to digital. The number of input and output channels required depends on the number of axes controlled. The on-board microprocessor unit is necessary to perform real-time calculations without burdening the host computer.

Another important consideration is the software of the motion control system. One of the most important questions to ask is what type of control algorithms can be implemented easily. Most of them support a linear algorithm known as PID control, containing the proportional (P), integral (I), and the derivative (D) of the feedback terms, and their combinations. However, in some cases, increasingly complex algorithms may be implemented to meet the required motion quality given a certain mechanical assembly. For example, adding another integral term to the PID terms may be necessary to sufficiently obtain a

constant velocity in a stiff flexural bearing stage. In such cases, having the "right" software can save days or even weeks of programming effort, which is certainly an important factor in prototyping.

References

1. A. H. Slocum, *Precision Machine Design*, Society of Manufacturing Engineers, Michigan, pp. 401–625 (1992).
2. L. Saggre, S. Kota, S. B. Crary, "A new design for suspension of linear microactuators," *Dynamic Systems and Control: Volume 2*, ADME DSC-Vol. 55-2, pp. 671–675 (1994).
3. Melles Griot Catalog, pp. 44–10, Melles Griot, Carlsbad, CA (1999).
4. P. Schellekens, J. S. E. Van De Pasch, and A. C. H. Van Der Wolf, "Design and results of new interference refractometer based on a commercially available laser interferometer," *Annals of the CIRP*, **35**(1), pp. 387–391 (1986).

Chapter 6
Interferogram Data Sampling

The analog interference signal recorded by the detector needs to be captured and digitized at the appropriate and predetermined OPDs. In this chapter, the instrument's subsystem that produces the digital pulses to "clock" the data acquisition of the interferogram is discussed. This subsystem is referred to as the *clocking system* or the interferogram clocking system throughout the book.

Figure 6.1(a) shows a simple schematic of the clocking system. Figure 6.1(b) illustrates the interferogram sampling process, where the digital pulses synchronized with the moving mirror's position are used to clock the data acquisition. The figure shows the most commonly employed scheme in which the clock pulses are separated by an equidistant interval, d_s. It also shows clock positional error, Δd_s, which is discussed in detail later in the chapter.

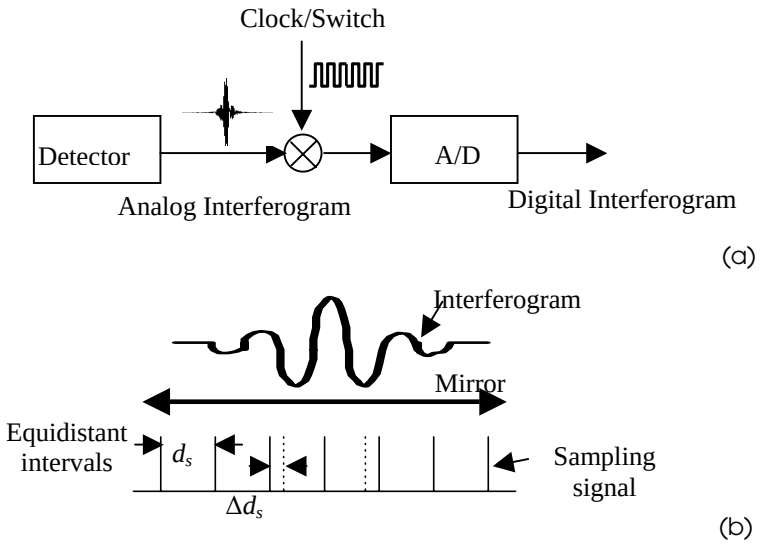


Figure 6.1 (a) Simple schematic of the interferogram data acquisition clocking system. (b) Illustration of interferogram data sampling. Digital pulses are usually used to clock the data acquisition such that their timing is synchronized with the moving mirror's position.

6.1 Step Scan vs. Continuous Scan

There are two main sampling schemes: *step-scan* and *continuous-scan*. Under a step-scan operation, the moving mirror advances to a new location and stops before the data is acquired. This is repeated many times until it reaches its end-point to meet the spectral resolution requirement. Under a continuous-scan operation, the moving mirror travels at a constant velocity while the data are acquired. The mirror is usually commanded to move back and forth and data from the same locations are summed and averaged to improve the SNR. The illustrations of both schemes are shown in Fig. 6.2.

For measurements that rely on a time-dependent phenomenon, a step-scan operation is usually necessary.[1] In most other applications, continuous-scan operations using "rapid" mirror motion is advantageous. Table 6.1 summarizes the general advantages and disadvantages of both methods.

There are three main requirements of the interferogram clocking system: (1) a small enough sampling period or interval, (2) high accuracy with respect to the optical path length difference (OPD), and (3) high repeatability between scans.

In the following sections, the effects of these factors to the output spectrum are discussed.

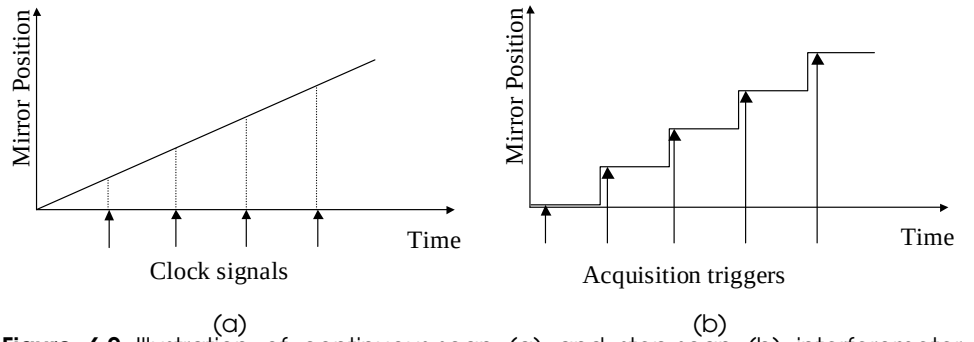


Figure 6.2 Illustration of continuous-scan (a) and step-scan (b) interferometer operations.

Table 6.1 Advantages and disadvantages of step-scan and a continuous-scan interferometer operations.

Step-scan	Continuous scan
The dc signal needs to be modulated to improve SNR. A mechanical chopper and its electronics need to be used before bandwidth-narrowing techniques (filtering) can be applied.	Signal is already modulated, with a range of frequencies that is determined by the speed of the moving mirror and the radiation wavelengths. Thus, bandwidth-narrowing techniques can be applied directly.
The length of the measurement time proportionally increases with the sampling frequency (highest wave number analyzed) and the retardation distance (resolution intended).	Generally, a faster measurement time; although the length of the measurement time depends on the retardation distance required, the sampling frequency requirement does not affect it.

Table 6.2 Continued.

Step-scan	Continuous scan
Allows the use of a mechanical chopper and lock-in amplifier detection, which can provide a higher degree of noise rejection than filtering alone.	Filtering techniques, such as Fourier-filtering limits the bandwidth detected, but does not differentiate the noise and the signal based on their phases.
There is no sampling error due to the electrical delays and the detector's limited bandwidth. Sampling errors due to sensor errors, however, still exist.	Sampling errors can arise due to the electrical delays and the detector's limited bandwidth; especially a concern in rapid-scanning interferometer

6.2 Sampling Period

6.2.1 The absolute minimum: the Nyquist criterion

To properly digitize a sinusoidal signal, the minimum sampling frequency f_s needs to be at least twice the frequency of the signal, known as the *Nyquist criterion*, otherwise *aliasing* occurs. Aliasing is a phenomenon where an analog signal is represented by a lower-frequency digital signal due to insufficient sampling frequency. The high-frequency signal is said to be *folded* into a lower frequency. This phenomenon is illustrated in Fig. 6.3. Clearly, applying the Fourier-transform routine to the digital signal, represented by the dashed line in the figure, results in erroneous frequency and spectral information.

For a broadband signal, the highest-frequency component (the shortest wavelength) determines the Nyquist frequency. It is known that for a given wavelength λ_0, the constructive interference occurs at every $\lambda_0/2$ of the mirror retardation distance. Thus, for radiation that contains the minimum wavelength λ_{min}, the fastest modulating signal in the interferogram has a period of $\lambda_{min}/2$, which is equal to $1/2\upsilon_{max}$. Therefore, according to the Nyquist criterion, the sampling period d_s should be

$$d_s \le \frac{\lambda_{min}}{4} = \frac{1}{4\upsilon_{max}}. \tag{6.1}$$

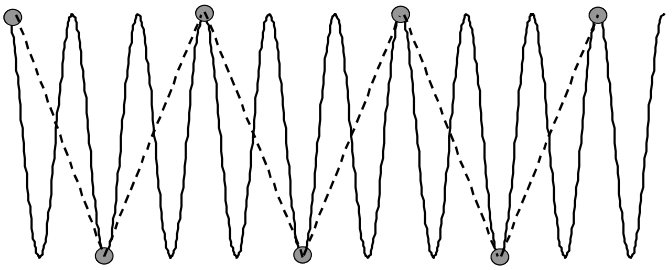


Figure 6.3 Analog signal (solid line) is sampled at insufficient sampling frequency (bullet points), resulting in erroneous computed digital signal (dashed line).

As an example, consider an FT spectrometer to produce a spectrum of visible light with a maximum wave number of 20,000 cm^{-1} (500 nm). In this case, the sampling distance should be smaller than 125 nm.

So far, the "sampling period" in the OPD domain has been discussed. However, in a continuous-scan operation, the sampling frequency in the time domain also needs to be considered. For mirror velocity v, the maximum frequency present in the interferogram is given by

$$f_{\max} = 2v\upsilon_{\max} . \tag{6.2}$$

It is important that the detector and its supporting electronics have a bandwidth that is substantially higher than $f_{\max}$. Following the Nyquist criterion, the sampling frequency in the time domain is then

$$f_s \geq 2 f_{\max} = 4v\upsilon_{\max} . \tag{6.3}$$

6.2.2 More than the Nyquist minimum

Practically, one should design the clocking system with a frequency greater than the Nyquist frequency. A factor of more than three or four times the Nyquist criteria is a good rule of thumb. Although the interferogram's frequency information is correct upon satisfying the Nyquist criteria, the accuracy of the magnitude information may be less than desirable. Figure 6.4 illustrates this problem where the magnitude of the digitized signal may not accurately represent the magnitude of the real signal when the "bare minimum" Nyquist frequency is used although the frequency information is correct.

Along the same line, also consider an interferogram containing two frequency components that are equal in magnitude. If sampled with a frequency that is just slightly higher than the Nyquist frequency, the two "peaks" in the spectrum have different magnitudes. Figure 6.5 shows two such peaks that come from two monochromatic signals with the same magnitude and frequency components close to the Nyquist sampling frequency. The result of real broadband or quasi-monochromatic radiation consisting of such frequency components is a "sheared" or distorted spectral shape.

In addition, one can also see from the illustration in Fig. 6.4 that slight changes in the clock positions greatly vary the resulting spectral magnitude. In fact, if the clock signals happen to be exactly at the zero-crossings, the recorded magnitude is zero. In other words, low-frequency sampling results in a spectrum that is more sensitive to sampling error.

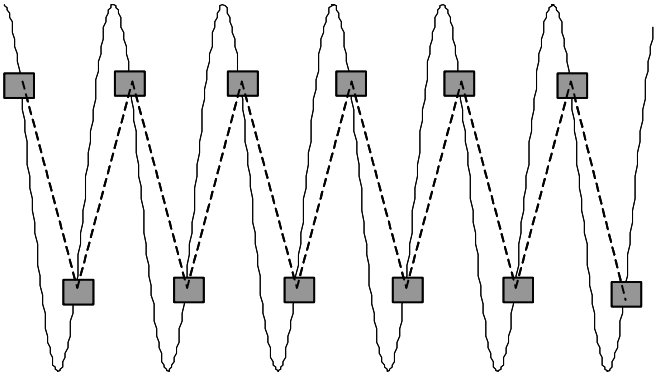


Figure 6.4 Degradation of magnitude accuracy due to limited sampling frequency. Solid line: the original analog signal. Dashed line: the digital representation.

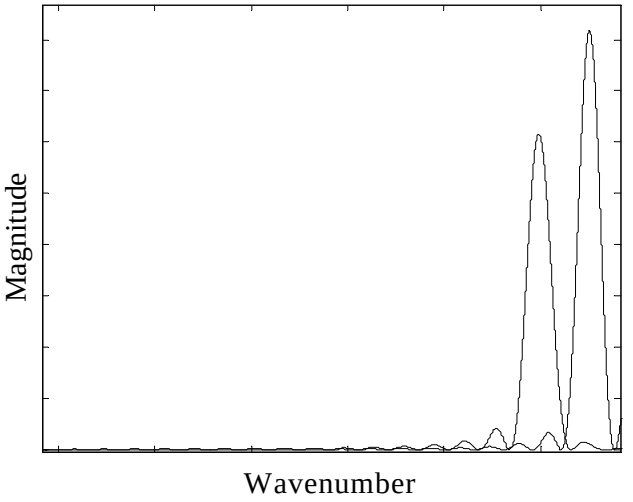


Figure 6.5 Erroneous spectral magnitude information resulting from low-sampling frequency. The two peaks come from monochromatic signals having the same magnitude with frequency components close to the Nyquist frequency.

6.2.3 Anti-alias filter

Real interferogram signals also contain some noise components. When designing the interferogram clocking system it is especially important to consider the high-frequency noise that can come from the detector electronics as well as the source or the background radiation. As the real signal can fold, so can the noise signal. Unlike its real counterpart, however, the noise signal is often unpredictable and its frequency components are usually unknown.

To avoid folding of the high-frequency noise (into the measured bandwidth), an anti-alias filter should be employed. An anti-alias filter is an analog filter

specially designed to attenuate the higher frequency noise by limiting the bandwidth of the measured interferogram. The simplest form would be a first-order RC filter, which is usually designed to have a cut-off at the Nyquist frequency. This filter is placed directly after the detector's preamplifier circuit, before the data acquisition. It should be designed to have a bandwidth high enough so as not to greatly affect the phase of the interferogram. The example below illustrates the idea.

Anti-alias filter design example:

$\upsilon_{max} = 1000\ cm^{-1}$ (given)
$v = 0.25$ cm/sec (given)
$\geq f_{max} = 500$ Hz
$\geq f_{Nyquist} = 1$ kHz
$f_{sampling} = 2$ kHz (designed to be twice the Nyquist frequency)
Type of anti-alias filter: first-order filter with 1-kHz cut-off frequency (at $f_{Nyquist}$)

Figure 6.6 shows the *Bode1* plot (frequency response plot) of the anti-alias filter. As seen on the figure, at 500 Hz, the phase lag introduced by the anti-alias filter is ~25 deg. This corresponds to ~25/360 deg of one cycle period of the sinusoid, which corresponds to about 0.14 msec (one cycle has a period of 1/500 second for a 500 Hz signal). This delay can be detrimental to the spectral SNR if the mirror velocity is not held perfectly constant. This matter is discussed in detail later in the chapter.

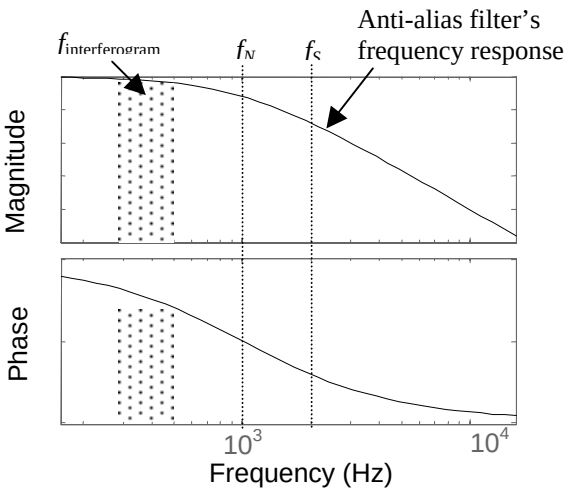


Figure 6.6 Bode plot of anti-alias filter. f_N: Nyquist frequency; f_s: sampling frequency.

1 Bode plots are log-magnitude and phase-frequency response curves as functions of log frequency.

6.3 Accuracy with Respect to OPD

Besides having a high enough frequency, the sampling-clock signal needs to be accurate with respect to the OPD. As discussed in the previous chapter, an HeNe-laser interferometer system is generally used to produce the sampling clock signal in an FTS, although other position sensing devices can also be used.

6.3.1 Effects of sampling errors to the output spectrum

In this section, the effects of the sampling errors on the output spectrum are analyzed. It is ideal to quantify these effects, for example, to predict the resulting spectral SNR for a certain magnitude of sampling error. However, such a task is rather difficult if not impossible. To produce a reasonably accurate SNR prediction due to sampling errors, one needs to take into account many interrelated factors, some of which are not easily quantifiable. These factors include characteristics of the error itself (e.g., random noise throughout the scan, localized jitters, or sinusoidal noise due to mechanical vibrations, etc.), location of the jitter with respect to the zero OPD, the sample's spectral shape, the maximum wave number, the number of spectral elements observed and the number of added and averaged interferograms, among others.

For a complex problem such as this, numerical simulation is a suitable approach. With numerical simulation, the analysis can be performed using errors with characteristics that best match the actual scenario, taking into account the complexities with less assumptions and simplifications. To guide the readers in that direction, a sample program that runs on Matlab and simulates a sampling error is provided in Appendix C. In the program, the interferogram signal and the sampling noise characteristics are input, and the spectrum is obtained.

Linear-sampling error: Using the program, examine the effects of a linear-sampling error (its magnitude being a linear function of mirror retardation distance, as illustrated in Fig. 6.7) to a spectrum that consists of monochromatic components. This type of error may be attributed to a steady change in air temperature, causing drift in the air refractive index, which in turns causes an erroneous reading from the laser interferometer. This is likely to be more prevalent in a step-scan operations, where the end-to-end scan can take up to a few minutes. Another possible scenario is one where an increasing angular error (with scan distance) is combined with offset distance between the laser interferometer's beam and the main interferometer's beam, usually referred to as an "Abbe error," as discussed throughout Chapter 4.

Figure 6.7 shows the sampling error as a function of the mirror retardation distance, and Fig. 6.8 shows its effects on the output spectrum. Specifically, Fig. 6.8 shows the error spectrum obtained by taking the difference between a spectrum produced without sampling error and a spectrum produced with the sampling error. In the simulation, three monochromatic signals at 5000 cm^{-1}, 10,000 cm^{-1} and 15,000 cm^{-1} are used. The total distance scanned is 1 mm, and

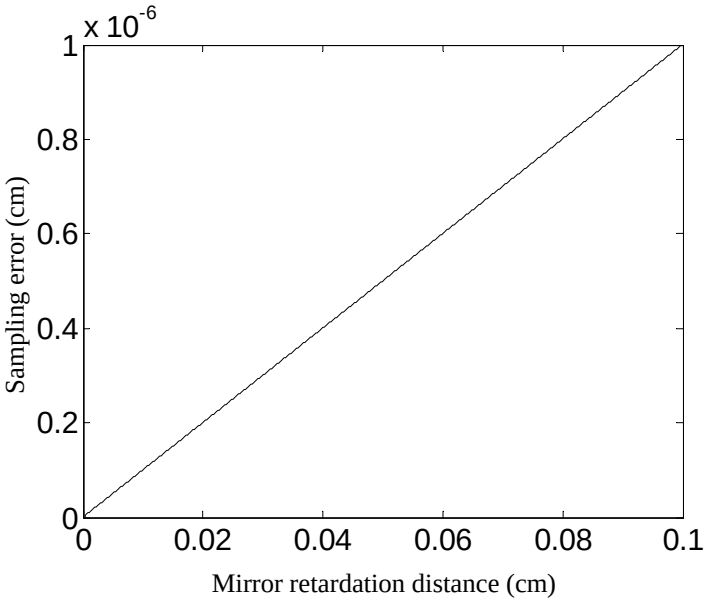


Figure 6.7 Linear-sampling error as a function of mirror retardation distance. Magnitude error at the end of mirror's travel is 10 nm.

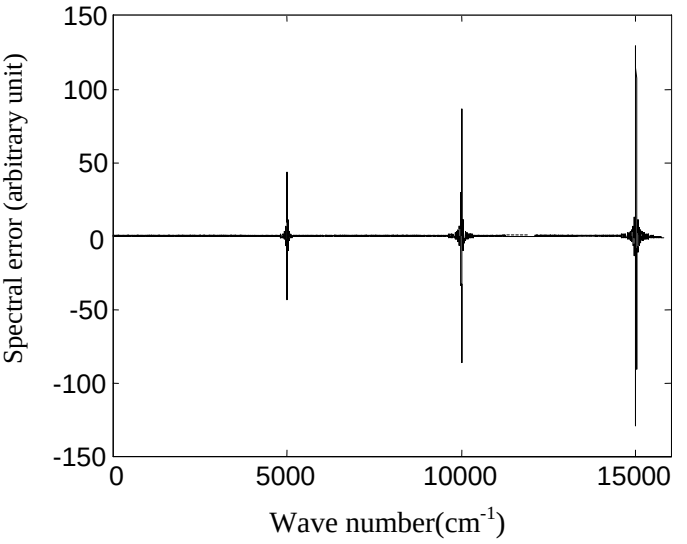


Figure 6.8 Error spectrum as a result of the linear-sampling error. This was obtained by taking the difference between the noiseless spectrum and the spectrum with the error.

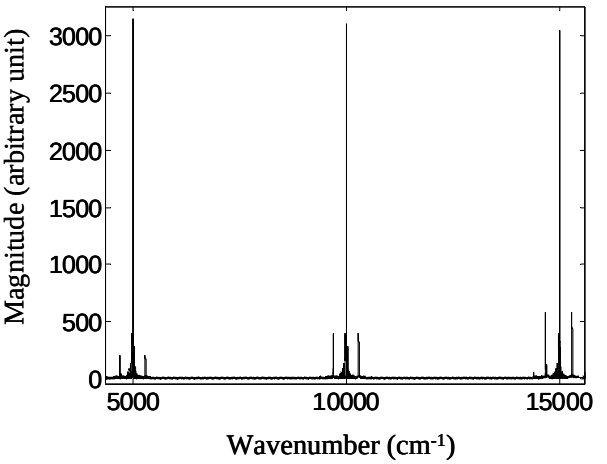


Figure 6.9 Output spectrum when sampling error is sinusoidal. Peak-to-peak error magnitude: 20 nm; frequency: 300 cycles/cm mirror travel.

the sampling period is 158.2 nm (approximately 1/4 of HeNe's 632.8-nm nominal wavelength). Further details of the parameters used can be seen in the appendix.

As indicated by the error spectrum, the high wave number components are affected more than the lower ones. As discussed in Sec. 6.2.2, this is a result of a "less than adequate" sampling frequency for the high wave number, making it more sensitive to sampling error. Although not apparent in the figure, the peaks in the spectrum are actually shifted slightly from their true locations. In essence, the wavelength information is no longer accurate. This is anticipated since the OPD information becomes erroneous due to the sampling error.

Vibrational-sampling error: The simulation program is also used to show the spectral error caused by sinusoidal-sampling error such as those caused by vibrations. In this case, the peak-to-peak magnitude of the sampling error is 20 nm and the frequency is 600 cycles/cm mirror travel. The resulting spectrum is shown in Fig. 6.9. It can be seen that secondary peaks on each side of the main spectral peaks appear. This behavior is similar to that resulting from sinusoidal-phase error, discussed in Sec. 4.2. Again, the higher-frequency components are more affected than the lower-frequency components.

6.3.2 Sources of sampling error

In Sec. 5.5, some of the sources of errors in the laser-interferometer-based position sensor were discussed. The analysis and results related to the positional-sensing errors are directly applicable to the sampling error, as the same signal is often used to produce the clock signal for the interferogram data sampling. These errors include Abbe error, sensor-alignment error, polarization error, and temperature drifts, among others.

In this section, another source of sampling error is examined: the combined effects of velocity variations and electrical delays. One would expect that when the signal from the position sensor is used to clock the data acquisition, the clock signal locations would be insensitive to the velocity variations of the moving mirror. In practice, however, because of the presence of electrical delays, this is not true. For an electrical delay t_{delay}, and a velocity error Δv, the sampling error Δd_s is given by

$$\Delta d_s = \Delta v \times t_{\text{delay}} . \tag{6.4}$$

The delays may originate from the finite bandwidth of the detector electronics and/or the anti-alias filter, as well as of the digital electronics.

Consider the numerical example given in Sec. 6.2.3, where the anti-alias filter causes an electrical delay of 0.14 ms for a mirror velocity of 2.5 mm/s. Assuming there is a velocity error of 2%, the sampling error is calculated to be 7 nm. For an application requiring high-spectral accuracy, this sampling error might be detrimental.

6.4 Repeatable Clock Position

Upon meeting the first two design criteria (the sampling frequency and the accuracy with respect to the OPD), the next task in designing the sampling system is to ensure that the clock position is repeatable from measurement to measurement and from scan to scan. In continuous-scan mode, a number of interferograms are often summed and averaged to improve the SNR. To achieve its purpose, however, the averaged data points need to come from the same OPD to a certain degree. Otherwise, the SNR may fail to improve.

As with the spectral effects due to sampling accuracy (Sec. 6.3.1), quantifying spectral SNR due to sampling repeatability is a difficult task. It depends on the shape of the transmitted spectrum itself, among various other things. Thus, numerical simulation is, in the author's opinion, the best way to reasonably compute and predict spectral SNR. The reader is welcome to use and modify or expand the program given in Appendix C to analyze the effects of sampling irreproducibility.

6.5 Hardware Considerations

As discussed before, a homodyne-laser interferometer system produces a sinusoidal signal with a period that corresponds to a distance traveled by the moving mirror, d_l, as specified by the following equation:

$$d_l = \frac{\lambda_{\text{laser}}}{2} . \tag{6.5}$$

Thus, if one uses a commonly employed HeNe laser with a wavelength of 632.8 nm, the distance between two zero-crossings is 316.4 nm. This simple

sampling scheme is adequate for FTS application in the infrared, or even in the long-wave near-infrared (> 2000 nm) region of the spectrum.

For applications in the visible or the short-wave near-infrared regions, one would need to design a more elaborate electronics system to produce a smaller clock period. The 316.4-nm sampling period can easily be halved by integrating the original sinusoid to create another sinusoid with a 90-deg phase lag. The electrical schematic for such design is shown in Appendix B. Because of the laser signal's ac coupling, variations caused by laser-power drifts and background-radiation noise, its accuracy is not affected. This sampling scheme has been used for FTS operation[2] in the near-infrared region between 1500 to 2500 nm, resulting in an SNR of ~10^4.

A method based on time-synchronization has also been used to clock data acquisition in continuous-scan operations. In this case, the mirror is commanded (and assumed) to travel at a very constant speed while data is being captured as illustrated in Fig. 6.10. Data acquisition is triggered when the interferogram reaches its peak at the zero retardation distance. Upon the trigger's occurrence, acquisition is timed by the internal clock of the data acquisition board.

Obviously, this method of sampling is very sensitive to velocity variations. It may only be appropriate for long wavelengths and low-resolution applications, where the external clocking capability is not available on the data-acquisition board. However, with the advent of nanometer and subnanometer precision positioning stages, one cannot simply denounce the technique. For the drive systems that do not use incremental position sensors, clock signals are not available without adding another piece of hardware. In those systems, time-based sampling schemes may be necessary.

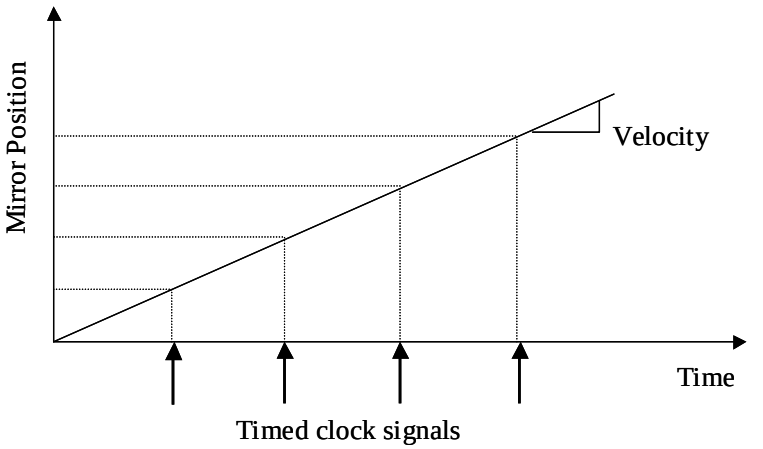


Figure 6.10 A sampling method in continuous-scan mode where the clock signal is produced by the internal clock of the data-acquisition system. Variations in the mirror's velocity result in accumulated sampling errors.

References

1. V.G. Gregoriou, M. Daun, M.W. Schauer, J.L. Chao, R.A. Palmer, "Modification of a research-grade FT-IR spectrometer for optional step-scan operation," *Applied Spectroscopy*, **47**(9), pp. 1311–1316 (1993).
2. V.A. Saptari, K. Youcef-Toumi, "Design and Performance evaluation of a near-infrared Fourier-transform spectrometer," *Proc. of SPIE*, **4163**, pp. 33–44 (2000).

Chapter 7
Data Acquisition

Data acquisition (DAQ) is a branch of the test and measurement field that functions to convert analog signals into digital data, and then streams the data into a computer for display and further digital processing. Today, there is a variety of off-the-shelf data-acquisition products for various applications characterized by the number of channels, the maximum speed, the resolution, etc.

This chapter briefly introduces the various DAQ formats and some of the factors to take into account when choosing one for FT spectroscopy applications. Of particular importance, however, is the last section of the chapter where the allowable spectral SNR due to the finite resolution of the analog-to-digital converters (ADCs) is analyzed.

7.1 DAQ Hardware Formats

The two main formats of DAQ hardware are internal (i.e., plug-in boards) and external. External systems have the benefits of being expandable (more channels, more signal-conditioning hardware possible, etc.) and having convenient connections. Thus, they are more suited to applications requiring a large number of inputs and outputs. Plug-in DAQ boards have direct data-bus connections; thus, they are capable of a high data-transfer rate. The data throughput of external devices is usually limited by the communication ports such as the serial, parallel, or the universal serial bus (USB) ports.

7.2 Analog Input

Some of the important considerations for analog inputs are the number of channels available, the maximum sampling rate, their resolution (in number of bits), their gain ranges (and whether they are software-selectable), and their accuracy. Each is discussed briefly in the following subsections.

7.2.1 Number of channels

Fourier-transform spectrometers require a minimum of one channel, which is needed for sampling the main interferogram. However, it is always convenient to have extra channels available for expandability. Additional channels are needed to implement compensation techniques such as light-source-drift compensation, temperature-variation compensation, etc.

7.2.2 Sampling rate

The maximum sampling rate, usually given in units of samples per second, determines the maximum amount of data that can be acquired in a given amount of time. It is important to recognize whether the specification refers to the maximum rate per channel or the maximum rate to be shared by the channels used. If one ADC is shared by a number of channels, the converter is *multiplexed**. In that case, the available sampling rate per channel is inversely proportional to the number of channels being used.

In addition to the maximum sampling rate, the settling time of the DAQ's amplifiers and relays needs to be considered in determining a "safe" sampling rate to use. For example, consider a data-acquisition board that has a maximum sampling rate of 1 MHz, four multiplexed analog input channels with a single ADC at 16-bit resolution. Theoretically, if all channels were used, the maximum sampling rate for each channel would be 250 kHz. However, depending on the amplifier and the delay of the electronics relay, the minimum time it takes to settle to within the 16-bit accuracy can be as large as 10–20 μs, resulting in a much slower sampling rate. It is obviously better to have as many ADCs as input channels, so that they are independent of each other and multiplexing is not needed.

7.2.3 Resolution

The resolution of ADCs is given in the number of available bits. For instance, 12 bits corresponds to 2^{12} levels, or 4096; similarly, 16 bits corresponds to 65,536 levels. The resolution of the ADC used to digitize the interferogram has a direct consequence on the achievable spectral SNR. This is discussed in detail in Sec. 7.4.

7.2.4 Input ranges

It is highly desirable to have DAQ hardware that has software-selectable gains, which allows for easy adjustments to the input-voltage range. For optimal use of

* Multiplexing refers to the method of switching between a number of input channels, when only one ADC is available.

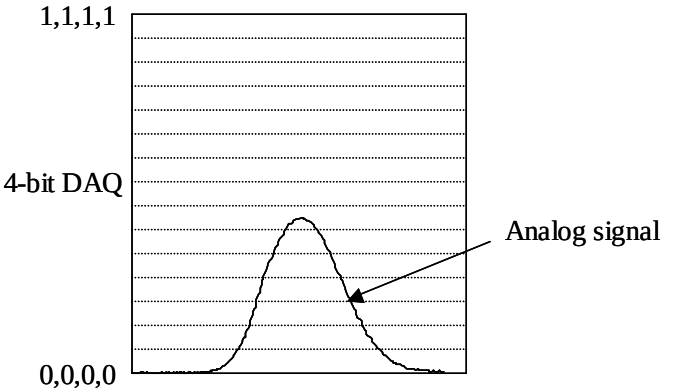


Figure 7.1 Nonoptimal use of a 4-bit DAQ's dynamic range.

the available resolution (number of bits), the DAQ's input-voltage range needs to be matched with the maximum analog-signal range. Figure 7.1 illustrates nonoptimal usage of the available A/D resolution where only half of the dynamic range is used. In FT spectroscopy, the maximum signal occurs at the zero OPD.

7.2.5 Accuracy

Ideally, when the digital number versus the input voltage is plotted, a perfectly straight line is obtained. Response nonlinearity and noise result in deviation from this transfer function. Accuracy is a measure of this deviation, which is given in units of voltage or least significant bit (LSB). The reader is advised to scrutinize the manufacturer's datasheet in regarding the accuracy of the DAQ system.

7.3 Hardware Triggering and Clock Signal

Interferogram data acquisition must be started (triggered) at a particular OPD, and clocked at certain OPD intervals and in real-time by signals directly corresponding to the mirror location, which is especially important in the continuous-scan operation. Therefore, the designer needs to make sure that the DAQ system has the external hardware triggering and clocking capability.

The clock signal usually comes from the interferogram of a reference laser, which would be a sinusoidal signal. This analog signal needs to be converted into digital pulses before it is suitable for the DAQ clock signal. The trigger signal, on the other hand, generally comes from an interferogram of a reference "white light" source that creates a centerburst at zero OPD. Using a *comparator* circuit, a digital pulse is produced when the interferogram exceeds a certain level, indicating that the mirror is at the zero-OPD point. Some DAQ systems have this comparator circuit already built in and the *trigger-level* software is selectable. Figure 7.2 illustrates the interferogram clocking and triggering process.

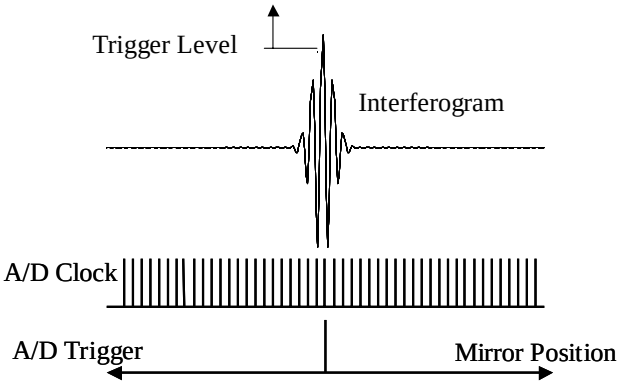


Figure 7.2 The interferogram clocking and triggering process.

7.4 Effects of ADC Resolution on Spectral SNR

The ADC used to digitize the amplitude of the analog interferogram has a finite resolution. It is desirable to know is how this affects the resulting spectral SNR in order to determine the ADC resolution to use. This problem can be considered in two slightly different ways. First, it can be assumed that the A/D conversion is perfectly repeatable. This means that a certain analog signal will be *rounded-off* to the same bit-number every time. The SNR as defined in Sec. 2.2 will be infinitely high. The problem then becomes that of the *detectivity*—that is, the minimum change required in the analog signal so it registers on the next bit-number.

The second way of looking at the finite-digitization-level problem is to assume that the A/D conversion has a random noise of ± 0.5 LSB. This means that there is uncertainty with a magnitude of 1-bit level (1 LSB) in converting a certain analog signal. Therefore, the SNR of the digital interferogram, and thus the spectral SNR, is determined by the total number of bits. Real ADCs have this type of noise or uncertainty. This process is continued until the allowable spectral SNR for a given A/D resolution is determined.

The SNR of the digital monochromatic interferogram obtained using an ADC with the number of bits b can be shown as

$$\mathrm{SNR}^{I}_{\mathrm{Digitization}} = \sqrt{3/2}k2^{b}, \tag{7.1}$$

where k is the portion of the fully dynamic range being used (a value between zero and one). The simulation result proving the validity of this equation is shown in Fig. 7.3. In the simulation, random noise with a magnitude of one LSB is added to the interferogram of a monochromatic signal. The SNR value is calculated by dividing the maximum interferogram signal by the noise rms value.

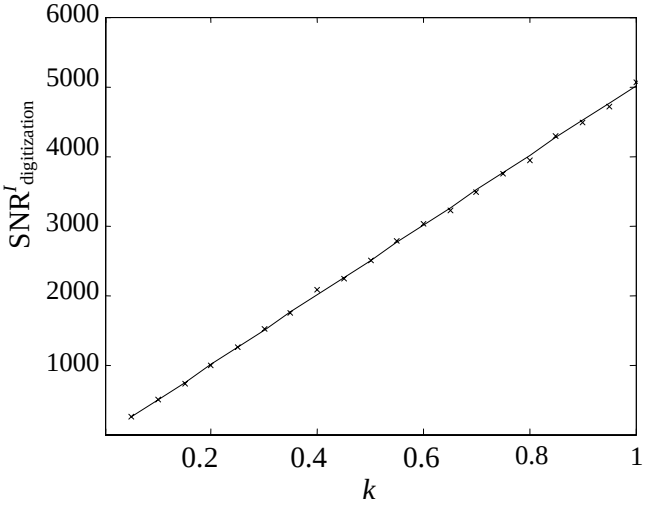


Figure 7.3 Interferogram SNR due to digitization noise as a function of k [Eq. (7.1)]. Solid line: theoretical value; x line: simulation result.

When going from the interferogram to the spectrum, the SNR increases by a factor of $N_I^{1/2}$, where N_I represents the number of interferogram points. The reasoning behind this factor is explained in the *multiplex advantage* discussion in Sec. 1.3.2. Thus, the spectral SNR is given by

$$\mathrm{SNR}^S{}_{\mathrm{Digitization}} = \mathrm{SNR}^I \times N_I^{1/2}. \tag{7.2}$$

It is now advantageous to examine the relationship between the number of interferogram points N_I and the number of spectral elements N. For a single-sided interferogram with a maximum mirror-retardation distance of $x_{\max}$ and a sampling period of d_s,

$$N_I = \frac{x_{\max}}{d_s}. \tag{7.3}$$

From Sec. 3.3.1 it is known that the maximum mirror-retardation distance determines the spectral resolution by the following equation:

$$\Delta\upsilon = \frac{1}{2x_{\max}}. \tag{7.4}$$

It is also known from Eq. (6.2) that the sampling period determines the maximum wave number:

$$\upsilon_{\max} = \frac{1}{4d_s}. \tag{7.5}$$

Note that the minimum wave number recorded in FTS is always zero because the dc signal is always taken into account in the Fourier transformation. Thus, the number of spectral elements N in the FTS is given by

$$N = \frac{\upsilon_{\max}}{\Delta\upsilon}. \tag{7.6}$$

Combining Eqs. (7.3)–(7.6), the number of the spectral elements N in a single-sided interferogram is related to the number of the interferogram points of N_I by

$$N_I = 2N. \tag{7.7}$$

Therefore, Eq. (7.2) can be rewritten as

$$\mathrm{SNR}^S{}_{\mathrm{Digitization}} = \mathrm{SNR}^I \times (2N)^{1/2}. \tag{7.8}$$

For a double-sided interferogram, it takes twice as many interferogram points to obtain the same spectral resolution. Thus, Eq. (7.7) becomes

$$N_I = 4N \tag{7.9}$$

for double-sided interferograms.

Broadband radiation can be thought of as a summation of monochromatic signals. Considering that the detector and the ADC have a finite maximum dynamic range, each "monochromatic band" occupies a portion of the available dynamic range. Thus, the signal of a particular band decreases in the presence of the other bands, as it has to "share" the available dynamic range. Taking this fact into account the spectral SNR for a monochromatic signal as described by Eq. (7.8) can be modified to

$$\mathrm{SNR}^S{}_{\mathrm{Digitization}} \approx \sqrt{2} \times \mathrm{SNR}^I \times N^{1/2} \times \frac{\rho(\upsilon)}{\rho_{\mathrm{total}}}, \tag{7.10}$$

for a broadband signal, where $\rho(\upsilon)/\rho_{\mathrm{total}}$ is the ratio of the spectral power density at the analyzed wave numbers to the total spectral power.

Now consider a numerical example—an interferogram of a broadband signal digitized using an ADC with a resolution of 2^{12} (a 12-bit converter). Assume that the full dynamic range is used, thus k in Eq. (7.1) is equal to 1. Also, assume that the maximum wave number is 5000 cm^{-1} as determined by the interferogram sampling frequency, and the resolution is 5 cm^{-1} as determined by the maximum OPD. Thus, the number of spectral elements N is equal to 1000. Finally, assume that $\rho(\upsilon)/\rho_{\mathrm{total}} = 0.2$. According to Eq. (7.1), the interferogram SNR is equal to $(3/2)^{1/2}(1)(2^{12}) = 5017$. Using Eq. (7.10), it is estimated that the allowable spectral SNR due to the finite resolution of the ADC is $(2)^{1/2}(5016)(1000^{1/2})(0.2) \approx 45{,}000$. Using a 14-bit ADC increases the SNR by a factor of 4, and using a 16-bit ADC increases it by a factor of 16.

Chapter 8
The Detector

Detector noise is probably the most studied and well-known source of error in spectroscopy, including Fourier-transform spectroscopy. However, contrary to popular belief, detector noise may not always be the limiting factor to the signal-to-noise ratio obtainable by a Fourier-transform spectrometer. It is the goal of this chapter to provide an understanding of the relationship between the magnitude of the detector noise and the spectral SNR allowable in FTS. In addition, various types of photo detectors and the general guidelines for selecting them are described.

It is important to understand that detector noise has a unique characteristic: unlike noise from other sources in the system, it is independent of the signal level. Thus, with the conventional definition of the signal-to-noise ratio,

$$\text{SNR} = \frac{\text{signal}}{\text{noise}}, \tag{8.1}$$

the SNR obtainable is proportional to the amount of signal available. What this means is that the SNR due to detector noise can be extremely high when the radiation energy that falls on the detector element is high. In this case, the actual instrumental SNR is limited by other components, which may be the mechanical drive error, the A/D digitization error, the sampling error, etc. Therefore, detector noise generally becomes significant only when the radiation power incident on the detector is very low.

8.1 Noise-Equivalent-Power

In order to quantify the effects of the detector noise on the output spectrum, one should consider noise equivalent power (NEP), a detector performance figure of merit that measures the minimum power within a specific bandwidth required to fall on the detector in order to produce a signal-to-noise ratio of 1. Thus, it gives a measure of the smallest detectable signal. The NEP has units of Watts/$\text{Hz}^{1/2}$. For a given bandwidth Δf in hertz, the noise power is

$$N_{\text{detector}} = \text{NEP} \times \Delta f^{1/2}. \tag{8.2}$$

This figure of merit is usually provided by the detector's manufacturer and it should include the noise contribution of the detector's preamplifier circuitry as well. For example, a thermoelectrically (TE)-cooled InGaAs detector made by Electro-Optical Systems, Inc. has an NEP value of $< 2.0 \times 18^{-12}$ W/Hz$^{1/2}$ at 25°C and $< 5 \times 10^{-13}$ W/Hz$^{1/2}$ at –30°C. Since the detector noise is independent of the rest of the system's conditions, the achievable SNR can be estimated by measuring or calculating the incident radiation power.

8.2 Spectral SNR Quantification

A practical method to estimate the spectral power, and thus the SNR in the detector-noise-limited case, is presented in this section. The method involves four main steps: (1) obtain the NEP value of the detection system; (2) obtain the radiation power information from the recorded interferogram; (3) calculate the SNR of the interferogram; and finally, (4) convert the SNR value of the interferogram into a spectral SNR value. Generally, the success of predicting the spectral SNR of the detector-noise-limited instruments hinges on the ability to quantify the source's radiation intensity. By measuring the actual radiation power that falls on the detector element, the need to calculate or estimate it is avoided. All of the other instrumental and procedural factors such as the optical throughput, the system's efficiency, and the sample absorbance are already taken into account as well.

Now the radiation power from the recorded interferogram is estimated. It is known that the interferogram of a broadband source consists of a "burst" at or near the zero OPD, which quickly drops as the OPD increases (see Fig. 8.1). The "wings" of the interferogram (the parts away from the centerburst) contain a signal that fluctuates around a mean value. This mean value is used to approximate the power of the radiation that falls on the detector. This part of the interferogram actually conveys the medium- to high-resolution elements of the spectrum, while the signal at zero OPD contains the dc and low-resolution elements. Therefore, it is reasonable to use this mean-signal value at the wings of the interferogram as the contributing signal in our SNR calculation.

For a photo detector using a transimpedance amplifier, the average radiation power at the wings of the interferogram can be approximated by

$$P_I = \frac{V}{gr}, \tag{8.3}$$

where V is the mean signal (in volts) of the interferogram wing, g is the transimpedance gain in volts per ampere, and r is the responsivity in amperes per

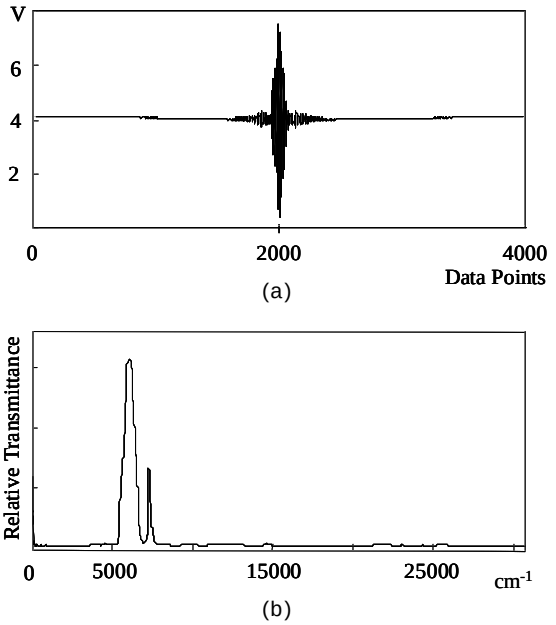
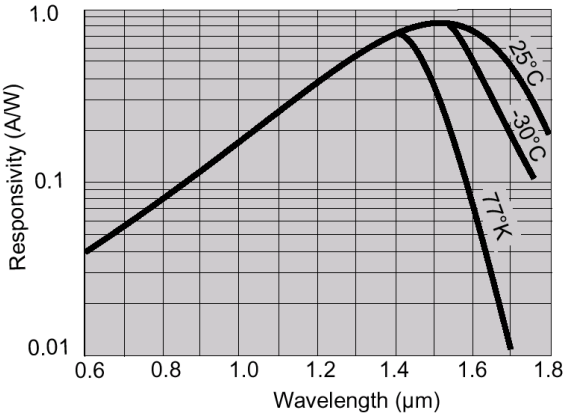


Figure 8.1 Interferogram (a) and its spectrum (b), obtained experimentally.

Figure 8.2 Responsivity curve for a germanium photodiode. (Reprinted by permission of Judson Technologies, Montgomeryville, PA.)



watt, which is a function of the wave number (see Fig. 8.2). For the current purpose of estimating the detected signal power, the average value of the responsivity over the analyzed wave number range suffices. With this information, the signal-to-noise ratio of the interferogram can be calculated as

$$\mathrm{SNR}^{I}_{\mathrm{detector}} = \frac{P_I}{N_{\mathrm{detector}}}. \tag{8.4}$$

The next task is to convert the SNR value of the interferogram to a spectral SNR value. As shown in Sec. 1.3.2, when going from the interferogram to the spectrum, the SNR is increased by the square root of the number of interferogram points N_I:

$$\mathrm{SNR}^S = \mathrm{SNR}^I \times N_I^{1/2}. \tag{8.5}$$

In addition, the spectral SNR depends on the power-density level of the analyzed spectral range compared to the total-power density. The higher the ratio, the higher the spectral SNR of that particular range analyzed (see Sec. 7.4). The spectral SNR is thus given by

$$\mathrm{SNR}^S \approx \mathrm{SNR}^I \times N_I^{1/2} \times \frac{\rho(\upsilon)}{\rho_{\mathrm{total}}}, \tag{8.6}$$

where the term $\rho(\upsilon)/\rho_{\mathrm{total}}$ represents the ratio of the spectral power of the analyzed range of wave numbers to the total power of the whole spectrum—the higher the ratio, the higher the SNR of the analyzed spectral range (see Sec. 7.4). This term actually signifies the importance of spectral bandwidth narrowing techniques, such as appropriate uses of filters to isolate the wavelength region of interest. This type of SNR improvement technique is discussed in Chapter 10.

It is important to emphasize that the quantification method presented here is meant to be an approximation, accurate for *order-of-magnitude* analysis. It is useful for error-budgeting purposes, where the user or the designer of the instrument would like to find out if the detector is the limiting component to the achievable spectral SNR. Although this SNR quantification method is not the only one and only viable method, it is hoped that the method presented here presents the reader with an intuitive procedure that he or she can use and modify as necessary.

Now a numerical example is considered, where the achievable spectral SNR is approximated while assuming the limiting factor is the detector noise. The system's variables are outlined in Table 8.1.

Table 8.1 System variables for numerical example.

Detector	
Type:	TE-cooled InGaAs
Spectral range:	800–2500 nm
NEP value:	0.5 pW/√Hz
Responsivity:	~1.5 A/W at 1700 nm
Transimpedance gain:	10^7 V/A
Bandwidth:	2 kHz
Spectrometer	
Type:	double-sided interferogram
Interferogram points:	2000 each side
Voltage level of wings:	~4 V
Analyzed range of wave numbers:	5850–6000 cm^{-1} with $\rho(\upsilon)/\rho_{\mathrm{total}}$ of 0.5

Figure 8.1 shows the interferogram and the corresponding spectrum used in the example. With a temporal bandwidth measurement of 2 kHz (without any bandwidth-narrowing technique), Eq. (8.2) calculates a detector-power-noise value of 22 pW. The average power of the radiation at the wings of the interferogram, using Eq. (8.3), is calculated to be $P_I = 4/(1.5 \times 10^7) = 267$ nW. Thus, with Eq. (8.4), the interferogram SNR is calculated to be $267 \times 10^3/22$, which is equal to ~12,000. With N_I equal to 4000 and $\rho_{analyzed}/\rho_{mean}$ equal to 0.5, Eq. (8.6) approximates the spectral SNR to be $\sim 3.8 \times 10^5$.

This SNR value is obviously quite high for a spectroscopic measurement. In cases like this, one must consider the other noise sources that may very well be the actual "performance limiter." However, if the experimental data proves that the spectral SNR within that particular wavelength range is indeed $\sim 10^5$, then it can be said that the FT spectrometer is operating at the detector's noise-limited condition.

8.3 Detector Types

There are mainly three types of detectors, and they are categorized based on their physical mechanisms in converting a radiation signal to an electrical signal: (1) photomultiplier tubes (PMT), (2) photodiodes, and (3) thermal detectors. A number of excellent books cover the various types of detectors and their operating principles in more detail than is presented here; the interested reader is referred to Refs. [1] and [2]. This section aims to give short descriptions on each type, intended to provide the reader with their characteristics and differences relevant to FT spectroscopic measurements.

Photomultiplier tubes are very sensitive and are fast radiation-energy detectors based on photoemissivity. A typical NEP rating for this type of detector is 7×10^{16} W/Hz$^{1/2}$ at 400 nm. A typical rise time is 1–2 ns. However, they are responsive only between 200 nm and 900 nm, making them useless for spectroscopic applications operating in the near infrared through the far infrared.

Photodiodes use semiconductor materials that produce current when radiation is absorbed. The responsive wavelength range depends on the detector material. Current semiconductor detectors cover any wavelength between 0.2 μm and 12 μm. Examples are silicon detectors that are responsive between 200 and 1100 nm, indium-gallium-arsenide detectors that are responsive between 800 and 2400 nm, and indium-antimonide detectors that are responsive between 2000 and 5500 nm. Their NEP values are generally on the order of 10^{11}–10^{13} W/Hz$^{1/2}$, a few orders of magnitude lower than that of a PMT. Because of their compact size, the extensive varieties that cover a wide spectral range, and their low NEP value, these solid-state detectors are the most widely used in spectroscopy today.

Heat-absorbing thermal detectors produce electrical signals from the temperature change in the detector element. Many of them have a flat spectral response between 0.2 and 20 μm. They are generally used for detecting higher radiation power levels, such as those in laser measurements because their power detection range is on the order of milliwatts through kilowatts. Their NEP value

is on the order of 10^{10} W/Hz$^{1/2}$. Detectivity and noise limitations often prevent them from being used for broadband, highly sensitive spectroscopic applications.

8.4 Selection Guidelines

There are many kinds of detectors commercially available today, varying in not only the type of material they are made, but also in their element size, cooling method, and type of electronics they possess. This is fortunate, but it also can be rather confusing to choose the right type of detector for a particular application. The detection system is relatively expensive, ranging anywhere between 10% and 60% of the total cost of the spectrometer. Thus, one cannot always afford to choose the "best" one out of those available. In fact, it is unwise to spend much of the budget on the detector while ignoring the performance capability of the other parts of the instrument. It is hoped that Sec. 8.2 provides the reader with enough knowledge to predict the spectral SNR allowable for a certain detector, so that the detection system is not under or over designed. This section aims to properly explain the important terminology used to specify a detector, although some of the terms have been mentioned in previous sections.

8.4.1 Spectral range and responsivity

The first consideration in selecting a detector should be its spectral range and its responsivity, $R(\lambda)$, usually given in amperes per watt. Responsivity is a measure of the detector's sensitivity to input radiation power, as given by

$$R(\lambda) = \frac{I_{\text{out}}}{P}, \tag{8.7}$$

where I_{out} is the detector output signal (in amperes) and P is the radiation input power in watts. The manufacturers usually provide a "responsivity curve" similar to the one shown in Fig. 8.2 for a germanium detector.

8.4.2 Active area

In certain cases, the active area of a detector determines the maximum throughput allowable. In the case of a detector-noise-limited system, a higher throughput results in a higher SNR. Thus, using a larger detector element is likely to improve the SNR, assuming that the throughput of the system is indeed limited by the detection-system optics rather than by the interferometer. However, larger active-area detectors generally have higher NEP value (higher noise), which may completely offset the SNR gain due to throughput increase.

8.4.3 Noise and detectivity

As mentioned in Sec. 8.1, the magnitude of the detector noise can be inferred from the NEP value as specified by the manufacturer. This NEP value should include both the detector noise and the noise of its supporting electronics such as the amplifier. If not, the amplifier noise should be considered separately and added to the detector-noise value.

Besides the NEP, D*, which denotes *detectivity*, is sometimes used to specify detector's noise performance. It is related to the NEP by

$$D^* = \frac{\sqrt{A_D}}{\text{NEP}}, \tag{8.8}$$

where A_D is the active area of the detector. D^* is usually given in units of $\text{cmHz}^{1/2}\text{W}^{-1}$. Thus, it is the inverse of the NEP value, normalized to the square root of a 1-cm^2 active area. The normalization facilitates a comparison between detectors of different active areas.

8.4.4 Linearity

Ideally, the output voltage of the amplifier is proportional to the radiation power. Otherwise, the problem of spectral accuracy and/or transferability between instruments arises. This is important in quantitative applications where, for example, the absorption magnitude of a particular solute is to be determined and compared with results or the standard measurement using other instruments.

Absorption is a measure of the ratio of the sample transmission T_s to the reference transmission T_r:

$$\text{Absorption} = f\left(\frac{T_s}{T_r}\right). \tag{8.9}$$

Figure 8.3 illustrates how absorption magnitude errors can occur due to detector nonlinearity. In the figure, P_s and P_r represent the transmitted radiation power through the sample and the reference, respectively. Because of the detector nonlinearity, an output-voltage error of ΔV_r occurs.

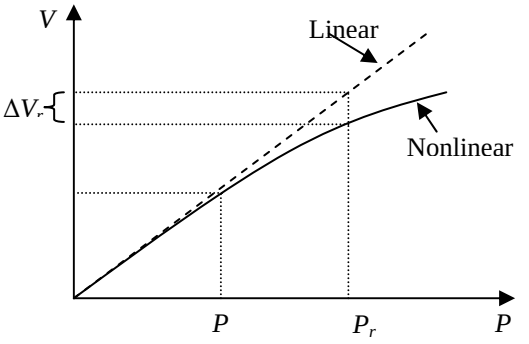


Figure 8.3 Illustration of how detector nonlinearity produces an erroneous absorption magnitude.

8.4.5 Temporal bandwidth

Section 6.2.1 explained that in continuous-scan operations, the modulation frequency of the interferogram is a function of the spectral components and the speed of the moving mirror. Rapid scanning is often advantageous as it enables a larger degree of data averaging in an allowable time, thus spectral SNR is increased. The bandwidth of the detection system for the continuous-scan operation needs to be carefully considered. Besides putting an upper limit on the scanning speed, the bandwidth of the finite detector and its electronics also creates a finite amount of delay. As discussed in Sec. 6.3.2, when this delay is combined with mirror-velocity errors, it can create significant sampling error that can ultimately reduce the allowable spectral SNR.

8.4.6 Summary

Table 8.2 shows a summary of InAs detector characteristics provided by the manufacturer Judson Technologies of Montgomeryville, PA. The table lists most of the main detector-performance parameters just discussed. Among them are the active area, the responsivity, the NEP, and the D* values. Most importantly, the table shows how the NEP value varies with the active area and the operating temperature.

Table 8.2 Sample of detector's specifications provided by most manufacturers. (Reprinted with permission of Judson Technologies, Montgomeryville, PA.)

Model Number	Part Number	Active Size (dia.) (mm)	Operating Temperature	Cutoff Wavelength @ λco (50%) (μm)	Responsivity @ λp (A/W)	Shunt Resistance R_D @ V_R = 10mV Min. (Ω)	Shunt Resistance R_D @ V_R = 10mV Typ. (Ω)	Maximum NEP @ λ_{peak} and 1KHz ($pW/Hz^{1/4}$)	Minimum D* @ λ_{peak} and 1KHz (Jones) ($cmHz^{1/2}W^{-1}$)	Capacitance C_D @ V_R = 0V (pF)	Optional Packages and Accessories
J12 Series Room Temperature InAs											
J12-18C-R250U	420002	0.25	22°C	3.60	1.5	200	300	6.0	3.7E9	50	LD2
J12-18C-R01M	420003	1.00			1.0	15	25	33	2.7E9	400	
J12-5AP-R02M	420011	2.00			0.8	5	10	71	2.5E9	1600	
J12TE1 Series One-Stage Thermoelectrically Cooled InAs											
J12TE1-37S-R250U	420088	0.25	-20°C	3.50	1.5	2000	3000	1.8	1.3E10	50	HS1, CM21
J12TE1-37S-R01M	420061	1.00			1.5	200	300	5.6	1.6E10	400	
J12TE1-37S-R02M	420065	2.00			1.25	50	90	13	1.3E10	1600	
J12TE2 Series Two-Stage Thermoelectrically Cooled InAs											
J12TE2-66D-R250U	420083	0.25	-40°C	3.45	1.5	12K	24K	.69	3.2E10	50	HS Amp, HS1, CM21, CM Amp
J12TE2-66D-R01M	420041	1.00				1.2K	2.4K	2.2	4.1E10	400	
J12TE2-66D-R02M	420089	2.00				300	500	4.4	4.1E10	1600	
J12TE3 Series Three-Stage Thermoelectrically Cooled InAs											
J12TE3-66D-R250U	420081	0.25	-65°C	3.40	1.5	160K	320K	.18	1.2E11	50	HS Amp, HS1, CM21, CM Amp
J12TE3-66D-R01M	420056	1.00				10K	20K	.71	1.2E11	400	
J12TE3-66D-R1.5M	420063	1.50				5K	10K	1.0	1.3E11	800	
J12TE3-66D-R02M	420098	2.00				2.5K	5K	1.4	1.2E11	1600	
J12TE4 Series Four-Stage Thermoelectrically Cooled InAs											
J12TE4-3CN-R250U		0.25	-85°C	3.30	1.5	400K	800K	.11	2.1E11	50	HS Amp, HS1, CM21, CM Amp
J12TE4-3CN-R01M	420093	1.00				25K	50K	.43	2.1E11	400	
J12TE4-3CN-R02M-B		2.00				6.5K	13K	.84	2.1E11	1600	

References

1. R. McCluney, *Introduction to Radiometry and Photometry*, pp. 59–66, Artech House, Inc., Norwood, MA(1994).
2. C. DeCusatis, *Handbook of Applied Photometry*, pp. 101–131, AIP Press, Woodbury, NY (1997).

Chapter 9
Consideration of Optics and Interferometer Alignment

In this chapter, the optical system's throughput for FT spectrometers is discussed first, and with the aim to help the designer select the proper size and "speed" of the focusing and the collimating optics, the diameter of the aperture, and the area of the detector element. Lenses and mirrors, the two types of focusing and collimating optics, are also briefly discussed. Finally, the alignment procedure for the interferometer is explained.

9.1 System Throughput

Optical throughput was defined in Sec. 1.3. It is revisited here and considered from an overall system perspective. In FT spectrometers, optical throughput is limited by either the interferometer's collimating optics or the detector's focusing optics.

In the detector subsystem, throughput is limited by the area of the detector and the "speed," quantified by the *f*/#[*] (*f*-number) of the detector optics. Generally, the light is collected using a fast detector lens (*f*/# ≈ 1) in order to "squeeze" the beam into the smallest detector possible because smaller detectors have better noise characteristics and are usually less expensive. The throughput of the detector optics can then be calculated to be

$$\Theta_D = A_D \times \frac{\pi}{4(f/\#_D)^2}. \tag{9.1}$$

The interferometer's collimating optics limits the throughput by limiting the aperture size, which is kept small enough to achieve the designed resolving power. In Sec. 3.3.2, the aperture-resolving power relation was calculated to be

[*] *f*/# (*f*-number) is defined as *focal length/beam diameter of a lens*—the smaller the number, the "faster" the optics.

$$a \le f\sqrt{\frac{1}{R}}, \tag{9.2}$$

where a is the radius of the aperture, f is the focal length of the collimating lens, and R is the resolving power. The throughput of the interferometer collimating optics can then be calculated to be

$$\Theta_I = \pi a^2 \times \frac{\pi}{4(f/\#_I)^2}, \tag{9.3}$$

assuming that a circular aperture is used. For the two subsystems to have the same throughput, the aperture should then be

$$a = \sqrt{\frac{A_D}{\pi}}\frac{f/\#_I}{f/\#_D} \tag{9.4}$$

by combining Eqs. (9.1), (9.2), and (9.3). Figure 9.1 shows the two constraints on the aperture radius as functions of R, and the following parameters are used to obtain the plot:

$f_{\text{collimating}}$ = 50 mm
$f/\#_I$ = 3
$f/\#_D$ = 1
detector diameter = 2 mm.

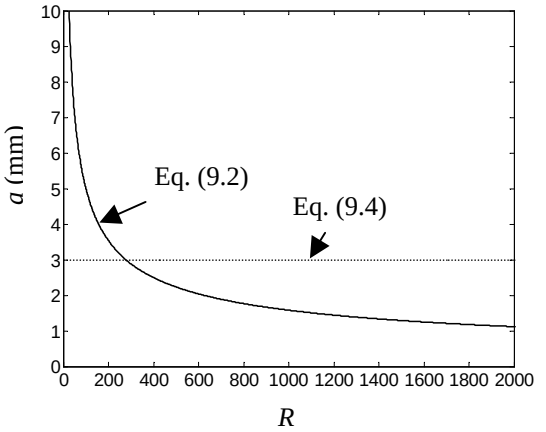


Figure 9.1 Constraints on the aperture radius imposed by the resolving power requirement. The straight line represents an aperture radius value when interferometer throughput is equal to detector throughput.

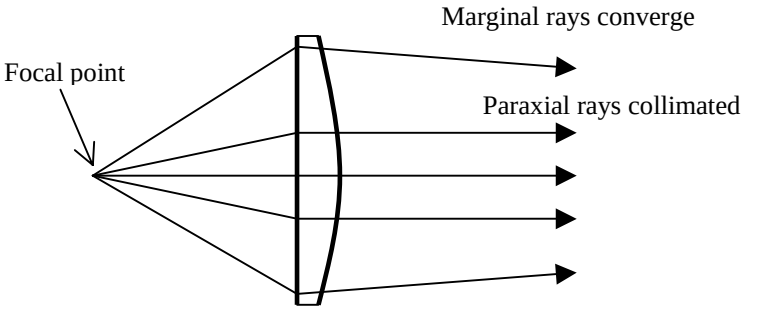


Figure 9.2 Illustration of the effects of spherical aberration in a plano-convex lens.

As a general rule, the throughput of a system should be limited by the detector optics for low-resolution applications (R < ~500). Thus, for low-resolution measurements, Eq. (9.4) should become

$$a \geq \sqrt{\frac{A_D}{\pi}} \frac{f/\#_I}{f/\#_D}. \tag{9.5}$$

Instruments designed for high-resolution applications, however, may have their throughput limited by the interferometer optics.

The upper limitation on the speed of the focusing and the collimating optics is usually due to optical aberrations. As the *f*/# is decreased, aberrations are increased and poorer beam quality is produced. For collimating optics, this means greater angle variations between the rays, while for focusing optics this means an increase in spot diameter. Figure 9.2 illustrates the effects of spherical aberrations in a plano-convex lens. To give a general idea, a single spherical lens with ~*f*/2 is generally acceptable for nonimaging focusing and/or collimating purposes.

9.2 Lenses vs. Mirrors

Many commercial FT spectrometers use off-axis parabolic mirrors for collimating and focusing the light. An important advantage of mirrors is that they have a much broader spectral range than the lenses used today. This is because the radiation does not have to go *through* the material, as in the case of lenses. For example, gold-coated mirrors reflect more than 98% of radiation between 0.7 and 10 μm. Another advantage of parabolic mirrors used as focusing or collimating optics is that they do not have chromatic aberrations where different wavelength components are focused at different points (see Fig. 9.3). Note, however, that they still possess other types of aberrations.

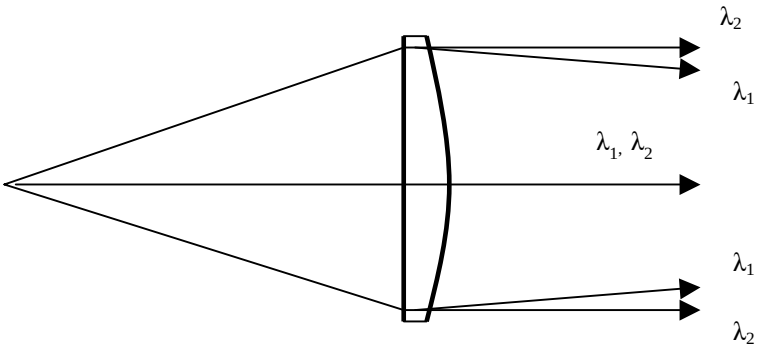


Figure 9.3 Illustration of chromatic aberration in singlet plano-convex lens ($\lambda_1 < \lambda_2$).

A disadvantage of using mirrors, however, is that they are harder to align, and usually result in bulkier systems than those using lenses. The beam needs to turn 90 deg for each reflection. The need for precise alignments also requires microadjustable optomechanical components. Their stability needs to be taken into the system's error budget.

9.3 Interferometer Alignment Procedure

This section outlines the general procedure to assemble and align an interferometer for broadband-light FT spectroscopy. Figure 9.4 illustrates the common setup, using lenses for beam collimation. The assumed highly divergent source requires two apertures. Aperture 1 (A1) is used to limit the area of the beam incident on *L*1, thus controlling the stray light, and aperture 2 (A2) is used to control the divergence angle of the beam entering the interferometer.

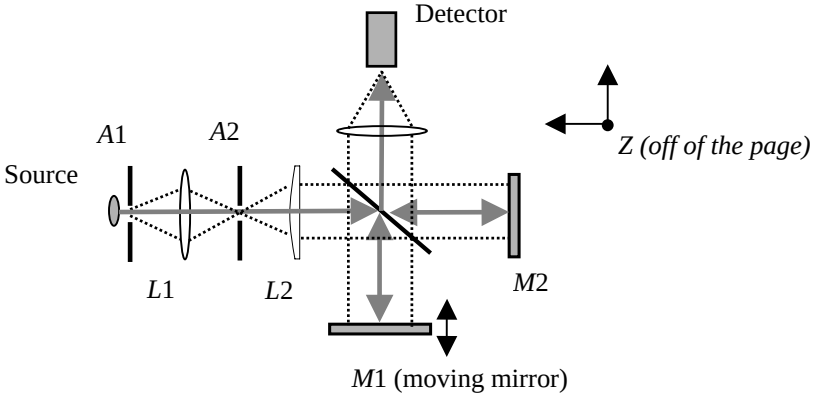


Figure 9.4 Common setup of a Michelson interferometer.

In the following alignment procedure discussions, it will be assumed that the moving mirror *M*1 only has one degree of freedom, which is in the linear *x*-axis. For ease and success of the alignment, *M*2 should have microadjustability in three directions: *yaw* (rotation about the *z*-axis, which is an axis perpendicular to the plane of the page), *pitch* (rotation about the *x*-axis), and in the linear *y* direction. Adjustability in the *y* direction is necessary when the movable mirror, *M*1, has a limited travel distance. Adjustment of the linear position of *M*2 enables modification to the position of the centerburst (at zero OPD) with respect to the absolute position of *M*1.

In addition, the beamsplitter needs angular adjustability around *x*, *y*, and *z*-axes. Note that after the alignment process is completed, both *M*2 and the beamsplitter should be mechanically locked into their final positions. In order to illustrate the complete alignment procedure, it is assumed that the interferometer is badly out of alignment, or that one is being assembled from scratch.

Alignment of M1 with respect to the optical axis. The first step is to position *M*1 so that it is perpendicular to the optical axis of the incident beam. In Fig. 9.4 this means that the mirror plane needs to be on the *y*-*z* plane, assuming that the *x*-axis is the optical axis. One way to do this is by adjusting the angular position of the beamsplitter about the *x*-, *y*-, and *z*-axes so that the beam is reflected directly back at the source. One should be able to see the reflected light imaged back on *A*1 (facing the interferometer). To ease the procedure, if *M*2 has been assembled, the beam reflected on *M*2 can be blocked (for example by using a piece of paper) so that there is an image only from *M*1. It is beneficial to check the angular errors of *M*1's drive system at this time by moving it back-and-forth throughout its travel range. The beam image on the aperture should not move and the position of the beamsplitter can be fixed.

Alignment of M2 with respect to M1. Now *M*2 is assembled, or the opaque material in front of it is removed to let the beam reflect on it. It is advantageous to remove the detector and replace it with a screen (a piece of white paper works), at a sufficient distance from the objective lens to get a clearly visible, enlarged image of the source. Now, the objective is to make the two image spots from the two mirrors overlap each other. This is done by adjusting the angular positions (yaw and pitch) of *M*2. To increase the precision of this procedure, a fine crosshair (perhaps made out of thin wires) can be placed on *A*1. The images of the crosshair on the screen should completely overlap.

Searching for the zero OPD. Now *M*2 needs to be positioned so that the location of the zero OPD relative to *M*1 is optimal. In order to maximize the achievable spectral resolution for a double-sided interferogram, the zero-OPD location should be near the center of *M*1's total travel distance, whereas for a single-sided interferogram, it should be close to the end of *M*1's total travel distance.

Mirror 1 needs to be locked into position (through servo control or by mechanical means) at the desired location. The detector is assembled, making sure that the image is focused on the active detector element. Using an analog or fast digital oscilloscope (> 10 MHz sampling rate), a signal is anticipated to

jump, indicating a centerburst, as *M*2 is slowly moved by its linear *y*-axis microadjustment. The adjustment must not cause significant vibration or angular motion of *M*2 or the centerburst may not come into view. For a broadband source, the jump of the signal would last a very short time (depending on how fast *M*2 moves and the bandwidth of the radiation). To lengthen this period, a narrow-band filter can be placed anywhere on the beam path (before or after the interferometer). This makes the signal burst longer, thus it is easier to "catch" and observe.

The final fine alignment. The final alignment can be done by moving *M*1 (under servo-control) back-and-forth, and maximizing the intensity of the centerburst signal. To do this, the angular position of M2 is adjusted and locked at a position where the maximum-centerburst intensity is reached. For a well-aligned interferometer, the maximum interferogram intensity should be close to twice the intensity at the "wings."

Chapter 10
Signal-to-Noise Ratio Enhancement Techniques

In this chapter, techniques to improve spectral SNR of measurements using an FT spectrometer are discussed. In attempting to improve spectral SNR, first it is important to recognize the limiting source of noise and error. In this chapter, some of the dominant sources of error are identified by analyzing the characteristics of the output spectra through practical and simple experimental procedures.

To begin, some of the various sources of errors are revisited, focusing mainly on their qualitative characteristics. Some practical ways to identify the instrument's limiting source of noise are presented. These procedures are certainly not set in stone, but are intended to provide the reader with a useful starting point in characterizing his or her instrument. Some accepted methods (for both hardware and software) to improve spectral SNR are then presented.

10.1 Identification of Error Sources

10.1.1 Detector noise

As discussed in Sec. 8.2, the allowable spectral SNR due to detector noise (in the detector-noise-limited case) is proportional to the radiation power P incident on the detector element, and is inversely proportional to the detector's NEP value:

$$\mathrm{SNR}_{\mathrm{detector}} \propto \frac{P}{\mathrm{NEP}}. \tag{10.1}$$

Therefore, detector noise is only a concern when the available radiation power is extremely low. By varying the signal power and noting the change in the spectral SNR obtained, one can check whether the instrument or the measurement is detector-noise limited. This can be done by placing neutral density filters in the

beam path or by varying the source intensity. If the recorded spectral SNR increases proportionally with the incident radiation power as suggested by

$$\Delta \mathrm{SNR}_{\mathrm{detector}} \propto \Delta P \,, \tag{10.2}$$

the detector noise would likely be the limiting source of noise. Otherwise, the limiting noise is likely to come from another component of the instrument.

In varying the incident radiation power, one must be careful not to affect other instrumental variables, such as the aperture size, the number of bits used, etc., resulting in an erroneous analysis caused by their interacting effects.

10.1.2 Digitization noise

The spectral SNR in the digitization-noise limited case is proportional to the available number of bits used to digitize the analog interferogram, as given by

$$\mathrm{SNR}_{\mathrm{digitization}} \propto 2^b k \,, \tag{10.3}$$

where b is the number of bits available and k is the bit-utilization efficiency factor (a value between 0 and 1). The detail of the analysis is given in Chapter 7. In contrast to the SNR due to detector noise, digitization-noise-limited SNR does not depend on the incident radiation power. One can vary the factor k and note the change in the SNR obtained to determine whether the instrument or measurement is limited by the resolution of the ADC. The factor k can be varied by changing the input range (or the gain) of the A/D input channel. Of course, one can also vary the radiation intensity, as long as detector noise does not play a role in affecting the SNR, that is, as long as the power is high enough to make the detector noise contribution negligible.

10.1.3 Interferometer alignment error

Interferometer dynamic-alignment error causes dynamic phase mismatch and, thus, the loss of spectral SNR, as analyzed in Chapter 4. Depending on the characteristics of the errors, different changes in the output spectrum can result. However, an important observation to note is that the spectral SNR or the magnitude of spectral deviations due to interferometer alignment error is proportional to the diameter of the beam (generally determined by the aperture diameter, D_{aperture}) in the interferometer:

$$\mathrm{SNR}_{\mathrm{interferometer}} \propto D_{\mathrm{aperture}} \,. \tag{10.4}$$

It is a more complex matter to determine whether the proportionality is linear because it depends on the exact beam profile and the frequency of the radiation present. However, this fact can be used to help us determine if our obtainable spectral SNR is limited by the interferometer dynamic-alignment error. In most commercial FT spectrometers, the diameter of the aperture is adjustable by software. Thus, it is a simple matter to examine the relationship between the

obtainable SNR and the aperture diameter. Of course, one must be careful of cross-dependence between the aperture diameter, the incident signal power, and the number of digitization bits used.

As seen in Fig. 4.10, if the alignment error is vibrational in nature, there would be secondary peaks or side bands when monochromatic radiation is analyzed. When the phase errors due to the dynamic-alignment error dominate, the magnitude of the ratio of the secondary peaks to the main peak is decreased as the beam diameter is reduced.

Another common alignment error is linear and varies proportionally with the mirror retardation distance. This type of alignment error can be easily identified by observing the recorded interferogram from a monochromatic source. A gradual decrease in the amplitude of the sinusoidal interferogram would be seen as the mirror travels.

10.1.4 Sampling error

The spectral noise due to vibrational sampling error is similar to that due to the vibrational interferometer alignment error; secondary peaks appear when the radiation is monochromatic. However, unlike the secondary peaks caused by the alignment error, those caused by the sampling error have magnitudes that are independent of the beam diameter.

Sampling error may be a result of the dynamic alignment error, the same source of error that generates phase error. In this case, it is difficult to determine which type of error is the dominant one. Since the magnitude of the effects of the sampling error to the output spectrum depends on the sampling frequency, as discussed in Sec. 6.2.2, the sampling frequency may be increased and the effects on the resulting SNR examined when it is suspected that our spectral SNR is limited by the sampling error. Of course, this is assuming that the hardware allows this to be done.

10.1.5 Light-source variation

Highly sensitive spectrometers do not matter much if the origin of the radiation is unstable or noisy. For applications that require extremely sensitive measurements, such as those requiring SNR on the order of $>10^5$, source-noise compensation techniques are likely to be required. One hardware technique is to use a double-beam method with which an additional detector is used to sense the source variation. This SNR improvement technique, among others, is discussed later in the chapter.

Light-source variation can be quantified easily by putting the detector directly in front of it and measuring the signal variation on an oscilloscope. For a source using an unregulated power supply, the dominant noise usually comes from the 60-Hz ac power line. Figure 10.1 shows measured fluctuations of a tungsten-halogen light powered by an unregulated dc power source. The effects

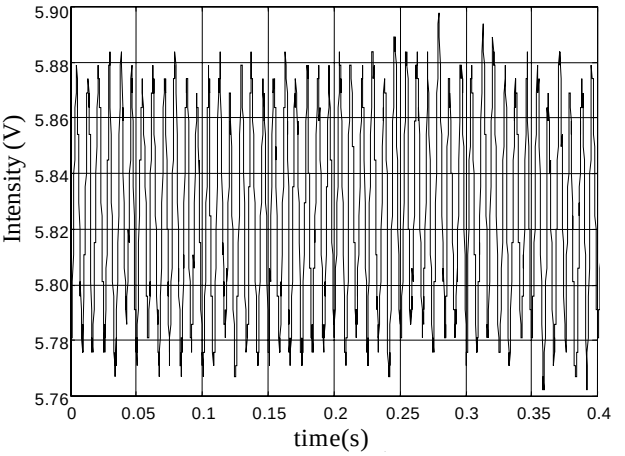


Figure 10.1 Power fluctuations of an unregulated tungsten-halogen light source.

of this type of noise, where it has well-defined frequency components, can easily be suppressed in FT spectroscopy. In a continuous-scan operation, the mirror speed can be chosen so that the signal modulation due to the wave-number range of interest lies outside the noise spectrum. Slow-varying drifts of the light source intensity are more detrimental to the spectral SNR, as they affect the scan-to-scan and measurement-to-measurement repeatability.

10.2 Averaging

If noise is random (i.e., it has uniform distribution over the frequency), signal averaging improves SNR. This is because coherent signals add more than the random noise signals. In continuous-scan operation, signal averaging involves scanning the mirror back and forth and averaging the interferogram data at each sampled position. For this technique to serve its purpose, the sampling error should be small. For a rapid scanning system, where phase delays introduced by the detector electronics are large, the interferograms from the forward and the backward scanning motion may be separated before they are averaged and Fourier transformed independently. This minimizes the difference of the origins of the added signals. In a step-scan operation, signal averaging involves integrating the signal at each sampled mirror location.

Assuming the signal is coherent and the noise is random, the SNR improvement can be explained by the following equation:

$$\mathrm{SNR} = \frac{S \times n}{N \times n^{1/2}} = \frac{S}{N} \times n^{1/2}, \tag{10.5}$$

where S and N are the signal and noise magnitude for each scan, respectively, and n represents the number of scans in a continuous-scan operation or the integration time in a step-scan operation.

10.3 Temporal Modulation and Bandwidth Narrowing

In a continuous-scan operation, the detector senses a modulated interferogram signal as a band of frequencies with limits given by the following equations:

$$f_{\min} = 2\upsilon_{\min} v, \tag{10.6}$$

and

$$f_{\max} = 2\upsilon_{\max} v, \tag{10.7}$$

where v is the mirror velocity in centimeters per second. Thus, the moving mirror velocity can be designed to create a modulated signal so that its frequency band is outside the frequency band(s) of the dominant noise. For example, if the 60-Hz source fluctuation noise is a problem, then modulating the interferogram signal in the audio frequency range minimizes its effect on the spectrum, as illustrated in Fig. 10.2.

In a step-scan operation, a mechanical chopper is generally used to modulate the signal at each sampled mirror location. The modulation frequency is thus determined by the rotational speed of the chopper.

Signal modulation does not improve the SNR value if it is not coupled with bandwidth-narrowing techniques, such as low-pass, high-pass, or band-pass filtering. These are called *temporal* techniques because they are applied in the time domain. Temporal bandwidth-narrowing techniques can be performed in both the analog and the digital domains (see Fig. 10.3). In the analog domain, electronic circuits are used to filter out or suppress unwanted signals with frequencies outside the interferogram frequency range. See Ref. [1] for excellent coverage on various analog filters, which includes useful circuit examples.

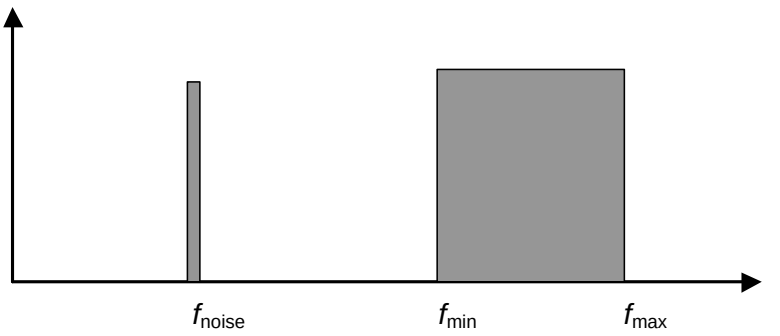


Figure 10.2 Illustration of frequency components and bands of noise and interferogram. Selecting the proper modulation frequency isolates the signal from the noise.

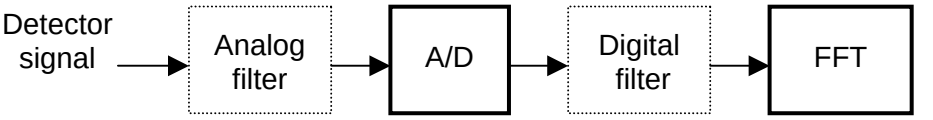


Figure 10.3 Temporal filtering in analog and digital domains.

Similarly, temporal filtration techniques can also be performed in the digital domain. This involves post-processing the data after it is acquired and stored in the computer. Of course, it is all done with software, thus faster implementation, modification, and optimization can be achieved. In addition, various nonlinear techniques can be implemented with great ease. The main drawbacks of digital processing are those due to the finite resolution (in both magnitude and time) inherent to A/D conversions. Noise in the megahertz range, for example, cannot be efficiently removed digitally from data that has been sampled at 100 Hz. In most cases, temporal filtering is done in both the analog and digital domains because they can complement each other.

10.4 Spectral-Bandwidth Narrowing

Spectral-bandwidth narrowing techniques aim to limit the detected radiation so that only the wavelength range(s) of interest is included. In doing so, the dynamic range of the detector and the ADC are efficiently used for the essential spectral features. Radiation in the spectral range that is of no value to the measurement is omitted and the available dynamic range is not "wasted." Mathematically the value of $\rho(\upsilon)/\rho_{total}$ that appears in Eqs. (7.10) and (8.6) is increased. Spectral-bandwidth narrowing makes sense only in the analog domain.

Spectral-selective filters (i.e., short-pass, long-pass, band-pass, and/or notch filters) are put in the beam's path to limit the detected radiation. They can be placed at any point between the source and the detector. As simple as this may sound, the selection and utilization of the appropriate spectral filters is an important practice that is often overlooked in experiments using FT spectroscopy. To illustrate the importance of this point, consider the spectrum shown in Fig. 10.4.

Assuming that this is a transmission spectrum of the buffer or the solvent used, evaluation of some absorption features of a particular solute in the buffer occurring somewhere in band *A* is of interest. Without any optical filter, most of the detector is filled, and the A/D dynamic range with signal energy comes from band *B*, which is not of interest for this measurement. Since $\rho(B)/\rho(A)$ is approximately equal to eight, almost an order of magnitude of improved SNR can be attained in this particular example just by selecting the appropriate wavelength range.

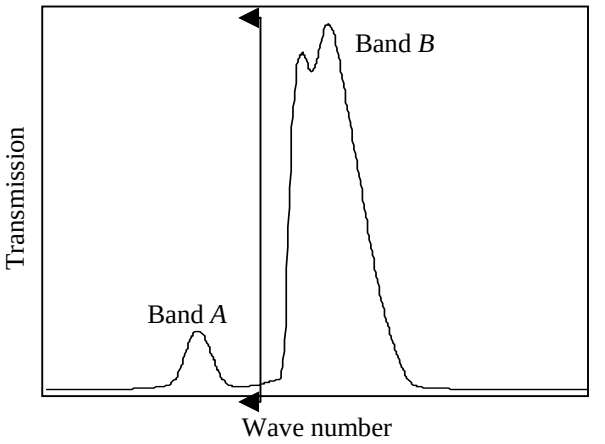


Figure 10.4 Spectral bandwidth-narrowing through long-pass filtering.

10.5 Spectral Post-Processing

So far, signal-processing techniques prior to spectrum generation (Fourier transformation of the interferogram) has been discussed. Upon spectrum generation, further data processing can be done to improve SNR. They can be performed on both the transmission or emission spectrum as well as the absorption spectrum. Familiar methods include baseline-variation removal and high-frequency noise removal. Note that the term high frequency here refers to the sharp spectral features in the spectral domain, rather than the time domain (Fig. 10.5).

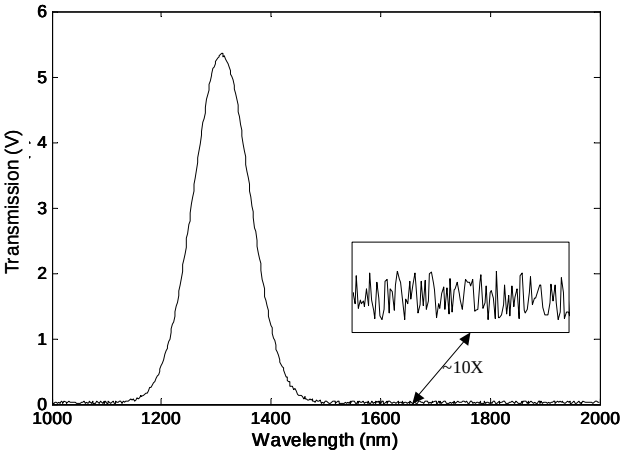


Figure 10.5 High-frequency spectral noise.

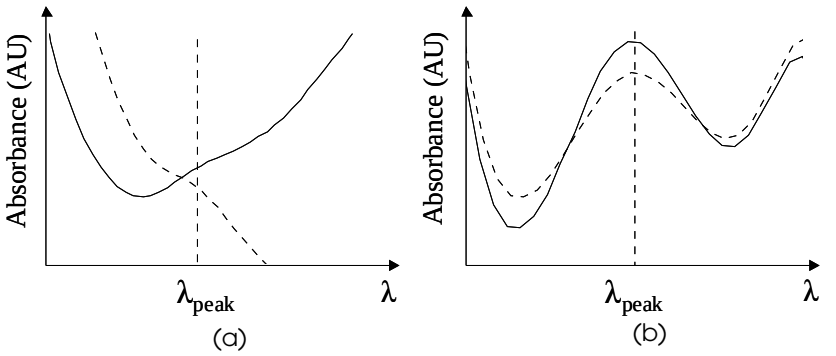


Figure 10.6 Spectral post-processing technique performed on the output raw spectrum.

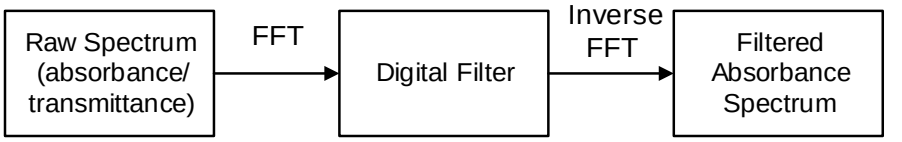


Figure 10.7 Illustration of the effectiveness of post-processing using a band-pass filter (Fourier filtering) to remove baseline variations. (a) Two raw-absorption spectra of an analyte in different concentrations of aqueous solutions. (b) The Fourier-filtered absorption spectra.

Methods to remove baseline variations include high-pass filtering and polynomial fitting. High-frequency spectral noise can be suppressed through low-pass filtering. If a certain spectral feature is expected, narrow band-pass filtering can significantly enhance the feature, and thus the SNR. This method of signal processing is sometimes called *Fourier filtering*,[2,3] and its process is illustrated in Fig. 10.6. Figure 10.7 shows the effectiveness of Fourier filtering using a Gaussian-shaped filter in enhancing spectral features buried under baseline variations.[4]

10.6 Double Beaming

The *double-beam* method calls for the source beam to be split in two with one focused on the sample and the other on a reference point. This technique is used when repeatability of the single-beam spectra is unsatisfactory due to source fluctuations, background noise, and even some instrumental or sample variations. Two detectors are then used to record the two transmission spectra simultaneously. This is called the *electrical subtraction* technique. Another type of double-beaming method involves subtraction in the optical domain. Both of these methods are discussed in this section.

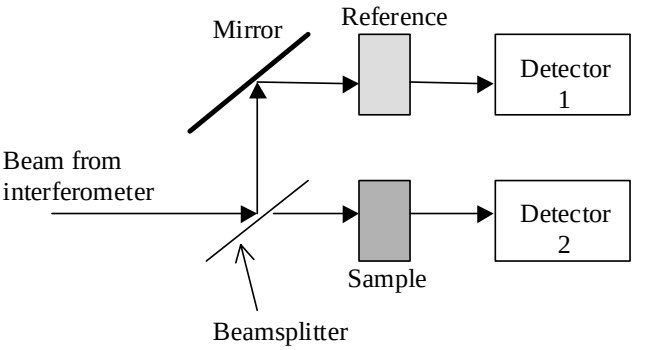


Figure 10.8 Simplified schematic of the electrical subtraction technique. Two detectors are used to record two interferograms simultaneously.

10.6.1 Electrical subtraction

Figure 10.8 shows a general architecture of the electrical subtraction technique. In this case, the beam is divided when leaving the interferometer and is detected by two detectors. The subtractions or other mathematical comparison techniques performed between the two interferogram signals can be done in the analog electrical domain or in the digital domain (prior to or after the A/D conversion). If done in the digital domain, two separate A/D channels are used. Signal comparison in the analog electrical domain may be advantageous because it may be used to improve the efficiency of the ADC's dynamic-range usage. Thus, it can improve the system's SNR if it is digitization-noise limited.

In practice, there are several potential problems to be aware of regarding double beaming. The first and foremost is the fact that no two detectors have the same dynamic characteristics. The difference in the phase response between them may distort the spectrum. Second, the noise and drift of each detector is not necessarily "in-phase" or synchronized with the other. This adds signal artifacts that may degrade the spectral SNR. In addition, one must also be careful with the detector's amplitude range and saturation limit. When subtraction is performed in the analog domain, detector saturation may not be easily monitored.

Changing the sample or reconfiguring the instrument's parameters may require readjustments of the detector-amplifier gain. Note that the electrical-subtraction technique does not improve results where spectral SNR is limited by the detector noise, nor does it improve the usage efficiency of the detector's dynamic range.

10.6.2 Optical subtraction

Rather than using two detectors to record the sample and the reference beam, the optical subtraction technique performs a signal comparison in the optical domain

prior to detection; therefore, it employs only a single detector element. This optical arrangement is also known as the *double-input-single-output* configuration. The benefit of this method is obvious: there are no errors introduced by the different characteristics of the two separate detector units. In addition, this subtraction technique optimizes the use of the detector's dynamic range. This is greatly beneficial in applications where the sample has a minute absorption signal, which causes poor use of the detector's dynamic range, as illustrated below

Consider an absorption measurement of a sample in an aqueous solution that absorbs only 1% of the incident radiation. In a single-beam measurement, the transmission signal recorded by the detector is T_b–A_s, where T_b is the transmission of the buffer (the pure aqueous solution) and A_s is the sample absorption signal. Thus, in this case most of the detector's dynamic range (99% of it) would be "wasted" in measuring the transmission signal of the buffer. Compare this scenario with a measurement using the optical-subtraction technique. The detector would only measure the magnitude of A_s, since the buffer transmission T_b has already been subtracted optically from the sample—buffer transmission T_b– A_s.

Retroreflector interferometers are generally used for double-beaming purposes. It allows two input and two output ports, as illustrated in Fig. 10.9. In the figure, it is configured for the optical-subtraction technique, where a single detector is used. A paper by Laroche et al. compares the electrical and the optical subtraction techniques through their implementations on emission measurements of infrared sources.[5] Another successful implementation of the optical-subtraction technique has recently been performed by P.S. Jensen and J. Bak.[6] It is used to detect the low-magnitude, near-infrared absorption of urea and glucose in aqueous solutions. The reader is also referred to a book by Griffiths for an excellent theoretical treatment of the subject.[7]

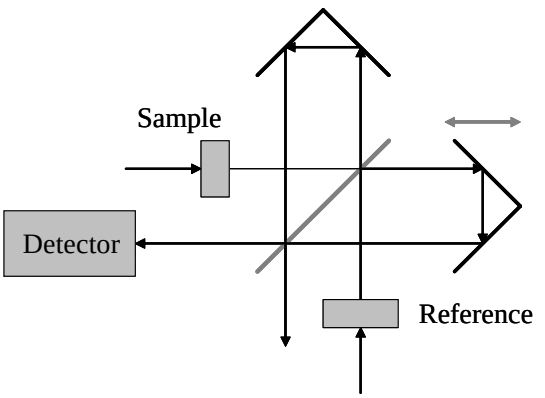


Figure 10.9 Optical subtraction technique using retroreflectors.

10.7 Gain Ranging

Digitization of an interferogram from a broadband source is an inherently inefficient process. At zero OPD, all the spectral components are added, resulting in a centerburst that quickly decays as the OPD increases. The generally more important medium- and high-resolution spectral information embedded in the wings of the interferogram uses only a small fraction of the available A/D's dynamic range or the number of bits. When spectral SNR is limited by the digitization noise (caused by an insufficient number of A/D bits), gain ranging may help improve it. This process involves changing the signal-amplification gain prior to digitization so that the analog signal fills the digital bits more efficiently throughout the interferogram. Of course, the different parts of the digital interferogram need to be normalized prior to the Fourier transformation. The reader is referred to Ref. [8] for a more detailed description of gain ranging.

10.8 Dynamic Tilt Compensation

As the name suggests, dynamic tilt compensation involves correcting the angular errors of the moving mirror in real time during the scan. For the three degrees of freedom: linear travel, pitch, and yaw corrections, three actuators and sensors (or sensing positions) are needed, as shown in Fig. 10.10. This compensation technique is now used in commercial FT spectrometers from several manufacturers, including Bomem (Quebec, Canada), Nicolet (Cambridge, MA, USA), Digilab (Randolph, MA, USA), and Shimadzu (Tokyo, Japan).

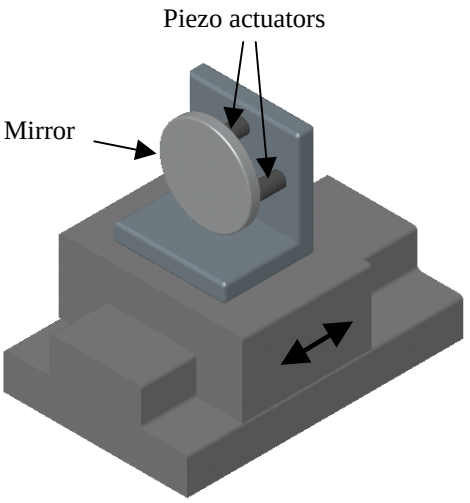


Figure 10.10 Possible mirror-drive design for dynamic alignment. Two piezo-actuators are used to perform fast tilt and yaw corrections.

The dynamic-tilt-compensation technique, or *dynamic alignment*, is greatly beneficial in UV and visible light applications where the short wavelengths require increasingly precise motion. Measurements that require a combination of high throughput (i.e., a large-beam diameter), high SNR, and high spectral resolution also benefit from this compensation technique.

References

1. P. Horowitz, W. Hill, *The Art of Electronics*, 2nd Ed., pp. 263–281, Cambridge University Press, New York (1989).
2. K. R. Betty, G. Horlick, "A simple and versatile Fourier domain digital filter," *Applied Spectroscopy*, **30**, pp. 23–27 (1976).
3. M. A. Arnold, G. W. Small, "Determination of physiological levels of glucose in an aqueous matrix with digitally filtered Fourier transform near-infrared spectra," *Analytical Chemistry*, **62**, pp. 1457–1464 (1990).
4. V. A. Saptari, K. Youcef-Toumi, "Sensivitity analysis of near infrared glucose absorption signals: toward noninvasive blood glucose sensing," *Proceedings of SPIE*, **4163**, pp 45–54 (2000).
5. G. Laroche, J. Giroux, A. Bordeleau, J. M. Garneau, "Investigation of electrical and optical subtractions using two input-port and two output-port FT-IR spectrometers," *Applied Spectroscopy*, **48**(3), pp. 356–362 (1994).
6. P. S. Jensen, J. Bak, "Measurements of urea and glucose in aqueous solutions with dual-beam near infrared Fourier transform spectroscopy," *Applied Spectroscopy*, **56**(12), pp. 1593–1599 (2002).
7. P. R. Griffiths, *Chemical Infrared Fourier Transform Spectoscopy*, pp. 171–184, John Wiley & Sons, Inc., New York (1975).
8. T. Hirschfeld, "Digitization noise in Fourier transform infrared spectroscopy and gain ranging," *Applied Spectroscopy*, **33**(5), pp. 525–527 (1979).

Appendix A
Simulation of Static-Tilt Error

```
%-------------------------------------------------------------------------
  %This program simulates phase-mismatch error due to static
    mirror tilt of a monochromatic signal
%-------------------------------------------------------------------------

    %Sampling Clock Signal Generation

    samp_period=1*632.8/4*10^-7;      %Sampling period in cm. (632.8 is the
                                      %wavelength of HeNe laser).
    scan_dist=0.5;                    %Mirror scan distance in cm.

    no_elements=floor(scan_dist/samp_period);
    x_samp=[0:samp_period:samp_period*no_elements]; %Vector of clock
                                                    %signals (in cm)

    %Phase-Shifted Interferogram Generation

    v=1000;                           %Wave number of the monochromatic
                                      % signal in /cm.
    tilt=0.0001;                      %Tilt angle in radians
    y=0;                              %Interferogram initialization
    no_loops=0;
    for x=- 1:.01:1                   %Cross-section of the beam
    phase=2*π*2*tilt*x*v;             %Phase-shift as a function of x
    y_int=cos(2*pi*2*v*x_samp+phase); %Interferogram with phase error
    y=y+y_int;
    no_loops=no_loops+1;
    end
    y=y/no_loops;                     %Normalization of interferogram
                                      %magnitude

%Zero fillings
```

```
total_data=2^14;
no_zero_fills=ceil[total_data-length(y)];
zero_fills=zeros(1, no_zero_fills);


%Triangular apodization
n_y=length(y);
points=[1:1:n_y];
tri=1-1/(n_y)*points(1:n_y);
y=y.*tri;                          %Dot product of interferogram with
                                   %triangular apodization function
y=[y zero_fills];                  %Zero-filling


%FFT operation
fft_y=fft(y);
fft_y(1)=[];
n_fft=length(fft_y);
spec_y=abs[fft_y(1:n_fft/2)];      %Spectrum generation

nyquist=1/(samp_period*4);
freq=(1:n_fft/2)/(n_fft/2)*nyquist;   %Frequency scale generation

plot(freq, spec_y);                %Plot of spectrum vs. wave number
```

Appendix B
Sampling Circuit Example

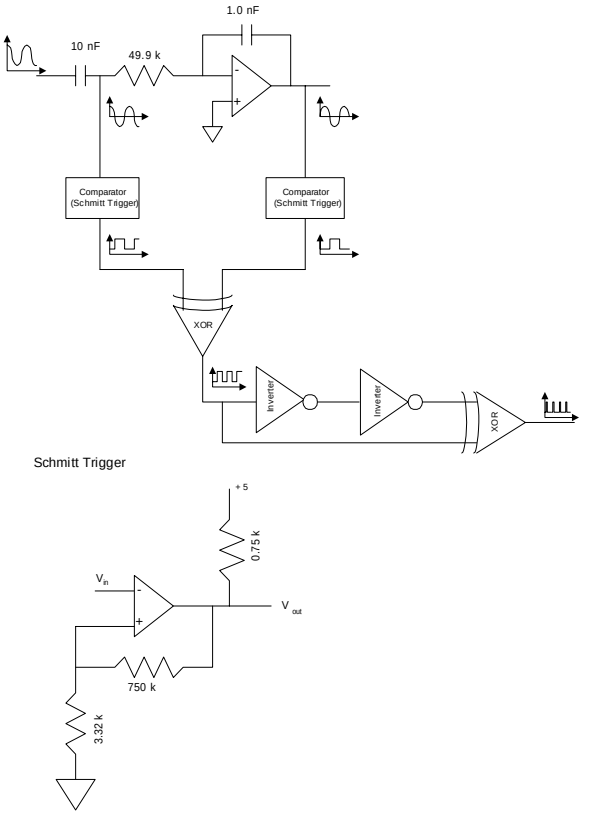


Figure B.1 This circuit converts a sinusoidal signal into digital pulses to clock the interferogram-sampling system. The sinusoidal input signal is first ac-coupled so that the zero-crossings are insensitive to laser-power fluctuation. The ac-coupled signal is then divided into two independent signals, and each goes through an integrator. Now, two separate channels exist, one a "cosine" and the other a "sine" (with a 90-deg phase difference between them). Each is then passed through a comparator circuit, consisting of a *Schmitt trigger*. The subsequent digital circuitry converts each of the rising and the falling edge of the Schmitt trigger's output into a digital pulse. Thus, a single sinusoidal signal is converted into digital pulses with a spacing of $\lambda/4$.

Appendix C
Simulation of Sampling Error

```
%-------------------------------------------------------------------------------------
    % This program simulates the effects of sampling errors to the resulting
     % spectrum.
%-------------------------------------------------------------------------------------

%Sampling Signal Generation

samp_period=1*632.8/4*10^-7;        %Sampling period in cm. is the
                                    %(632.8 wavelength of HeNe laser).
scan_dist=0.1;                      %Mirror scan distance in cm.

no_elements=floor(scan_dist/samp_period);
x_samp=[0:samp_period:samp_period*no_elements];
                                    %Vector of clock signals


%Sampling noise generation

%noise=10*1e-7*x_samp/max(x_samp);      %Linear error in cm
noise=20*1e-7*cos(2*pi*300*2*x_samp);   %Sinusoidal error in cm
%noise=1e-7*rand[1,length(x_samp)];     %random error in cm
%noise=1e-7*ones[1,length(x_samp)];     %constant error in cm

x_noisy=noise+x_samp;               %Vector of "noisy" sampling clock
                                    %signals

%Radiation signal generation
v1=5000;                            %monochromatic signal at 5000 cm^-1
v2=10000;                           %monochromatic signal at 10000 cm^-1
v3=15000;                           %monochromatic signal at 15000 cm^-1

arg=2*pi*2*x_samp;
```

```
argn=2*pi*2*x_noisy;
y=cos(arg*v1)+cos(arg*v2)+cos(arg*v3);          %Interferogram with no
                                                %sampling error.
y_n=cos(argn*v1)+cos(argn*v2)+cos(argn*v3); %Interferogram with
                                                %sampling error.

%Zero fillings
total_data=2^18;
no_zero_fills=[total_data-length(y)];
zero_fills=zeros(1, no_zero_fills);

%Triangular apodization
n_y=length(y);
points=[1:1:n_y];
tri=1-1/(n_y)*points(1:n_y);
y=y.*tri;                                       %Dot product of interferogram with
                                                % triangular apodization function
y=[y zero_fills];                               %Zero-filling
y_n=y_n.*tri;
y=[y zero_fills];

%FFT operation
fft_y=fft(y);
n_fft=length(fft_y);
fft_y(1)=[];
fft_yn=fft(y_n);
fft_yn(1)=[];

spec_y=abs[fft_y(1:n_fft/2)];                   %Spectrum generation
spec_yn=abs[fft_yn(1:n_fft/2)];                 %"Noisy" spectrum generation

nyquist=1/(samp_period*4);
freq=(1:n_fft/2)/(n_fft/2)*nyquist;             %Frequency scale generation

plot(freq, spec_y-spec_yn)                      %Plot of the difference between the
                                                %noiseless and the noisy spectra
```

Index

Abbe error, 58
aberrations, 93
absorbance unit (AU), 13
accuracy, 11
air bearing, 41
angular errors, 35, 44
aperture, 4, 5
apodization function, 27, 32, 33
Argand diagram, 21
averaging, 15

bandwidth, 11
beamsplitter, 17, 22–24, 34
bearing, 36, 41, 44

centerburst, 25, 34
collimating optics, 4
complex number, 21
constructive interference, 17
continuous scan, 46–47, 50, 53, 56

data clocking, 9, 43
degree of freedom, 35
destructive interference, 18
detectability, 15
detector linearity, 12
dispersive spectrometers, 1
disturbance, 41
divergence angle, 4

electromagnetic actuator, 46–48
electronic delays, 34
encoders, 53

feedback control, 45
Felgett advantage, 4, 6
filter spectrometers, 2
filter wheel, 2
flexural bearing, 41
friction, 46–51

gratings, 1

helium-neon lasers, 20
heterodyne interferometer, 52–56
homodyne interferometer, 53–58

instrumental line shape (ILS), 27
intensity, 21
interference function, 23–25
interferogram, 3–9
interferometer
  heterodyne, 52–56
  homodyne, 53–58

Jacquinot advantage, 4
Jacquinot stop, 29

linear errors, 35

Matlab, 37, 40, 41
Michelson interferometer, 17–18, 22–24, 31
mirror tilt, 26
mirror-scan distance, 31
modulation error, 35
monochromatic light, 17–29
monochromators, 1, 4

motion controller, 45, 48
multiplex advantage, 4, 6

optical path difference (OPD), 18
optical path length (OPL), 22

pitch, 35–37
phase
  difference, 18
  error, 35–37, 44
  mismatch, 40, 43
  shift, 39, 40
phasor, 21, 22
piezoelectric actuators, 46
polychromatic source, 24
position sensor, 45, 53, 56–61
prisms, 1
propagation number, 20

reference spectrum, 12
reflectance, 24
refractive index, 20, 34
repeatability, 11–12
resolving power, 4–5
retardation, 18–19, 25, 31
retroreflectors, 35–37
roll, 35–36
root mean square (rms), 12

sampling error, 35, 43–44
sinc function, 27, 32
secondary peaks, 41
sine error, 38
single beam, 12
sinusoidal signal, 18, 29
sinusoidal wave, 25, 32
slit, 1
solid angle, 4
spectral noise, 8
spectral resolution, 4–7
spectrographs, 1
spectrometers, 1, 2
  dispersive, 1
  filter, 2
spectral resolving power, 5
speed of light, 20
steradians, 4
step and settle, 47
step scan, 47– 56
stick-slip, 50
stiffness, 47– 59

temporal frequency, 20–21
throughput advantage, 4
tilting-filter, 2
top-hat, 27
transmittance, 24

wave equation, 22

yaw, 35–37

zero filling, 28, 33

Vidi Saptari is currently a Research and Development scientist/engineer at Delta Search Labs, Inc. (DSL) in Cambridge, Massachusetts. His expertise is in the design of high-precision electro-mechanical and optical systems. He has developed custom, highly sensitive Fourier-transform spectrometers for quantitative chemical measurements in the visible and near-infrared regions. One of his current projects is the development of a noninvasive, optical glucose sensor. Prior to joining DSL, he worked for Danaher Precision Systems in Salem, New Hampshire where he worked as a systems design engineer involved the design and manufacturing of one-of-a-kind nano- and micro-precision automation and robotic devices for the biotechnology and semiconductor markets. Mr. Saptari received his M.S. degree from the Mechanical Engineering department at Massachusetts Institute of Technology where he is also a Ph.D. candidate.