9

Introduction to Fourier Transform Spectroscopy

Julius Cohen

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards
Gaithersburg, MD 20899

March 1986

U.S. DEPARTMENT OF COMMERCE
)REAU OF STANDARDS

NBSIR 86-3339

INTRODUCTION TO FOURIER TRANSFORM SPECTROSCOPY

Julius Cohen

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards
Gaithersburg, MD 20899

March 1986

U.S. DEPARTMENT OF COMMERCE, Malcolm Baldrige, ***Secretary***
NATIONAL BUREAU OF STANDARDS, Ernest Ambler, ***Director***

INTRODUCTION TO FOURIER TRANSFORM SPECTROSCOPY

Julius Cohen

Radiometric Physics Division
Center for Radiation Research
National Measurement Laboratory
National Bureau of Standards
Gaithersburg, MD 20899

PREFACE

Fourier transform spectroscopy is being used extensively for chemical, optical, and astronomical studies, and owing to its great advantage in speed compared to conventional dispersive spectroscopy, it continues to grow in popularity. Other advantages may include higher resolution and convenience in experimentation.

In essence, Fourier transform spectroscopy results from the application of Fourier transform mathematics to interferometry. This report considers a simple Michelson interferometer only, which is an optical instrument for regulating optical path difference, thus producing a pattern of temporally and spatially varying light intensity termed the interferogram. The spectrum is educed from the interferogram by computations which are easily performed by a computer, but otherwise would be interminable. And although a large variety of interferometers are extant, the Michelson interferometer is the most widely used.

State-of-the-art Fourier transform spectrometers are so sophisticated and automated, that they may be operated by experimenters who understand virtually nothing about the principles of Fourier transform spectroscopy. This report is intended to be readable by scientists and program managers who have little or no prior background in Fourier transform mathematics or optics. Thus, first a review of rudimentary Fourier transform mathematics is presented, after which simple basic theories of the Fourier transform spectroscopy are developed. This approach is believed to be more digestible than giving mathematical fragments concurrent with the development of the theories of the spectroscopy, as is usually done in other texts. Another uncommon feature of this work is its copious use of graphics and its simple graphic solutions rather than the more involved mathematical derivations.

In a nutshell then, I have attempted here, without complications, to bring the novice to the point of understanding and appreciating that the Fourier transform of the interferogram is indeed the spectrum; in other words that seemingly unintelligible data, as shown in the example of Fig. 37, can truly yield accurate, distinct spectra. And in the interest of simplification the interferometer was idealized. However, the present report is not meant to be a substitute for a text or treatise on the subject, but rather to serve as a springboard.

Finally, it should be noted that this report is the first in a projected series of three. The second shall deal with practical aspects of Fourier transform spectroscopy--a detailed examination of a real, rather than idealized, Michelson interferometer--and the third shall be an assessment of sources of experimental error.

CONTENTS

List of Figures

Look to the essence of a thing.

Marcus Aurelius

One picture is worth more than ten thousand words.

Chinese proverb

INTRODUCTION TO FOURIER TRANSFORM SPECTROSCOPY

Julius Cohen

Radiometric Physics Division
Center for Radiation Research

This document is a simplified, concise introduction to Fourier transform spectroscopy. The emphasis is on elementary concepts and comprehension, and abundant diagrams are provided as an aid. The work is organized into three parts: first, a selective, but adequate review of Fourier transform mathematics, next, a treatment of the physics of a simple Michelson interferometer, and last, salient topics in Fourier transform spectroscopy.

Key Words: Apodization; coherence; convolution; deconvolution; Fourier transforms; infrared; interferometry; phase; resolution; sampling; spectrophotometry.

1. INTRODUCTION

Interferometers were already extant when, around the turn of the century, Albert Michelson (1852-1931) invented a much improved and versatile type which was capable of precise and simple regulation of the path difference between

---

[+]Interferogram is a trace of light intensity as a function of optical path difference; sometimes as a function of time. The spectrum, which is sought, is the light intensity as a function of wave number.

beams of light. It was realized that the Fourier transform of an interferogram, the pattern of interference fringes thus formed, gave the spectrum. There being no digital computers available at the time, Michelson constructed an analog computer which enabled him to carry out the computations, but the time required was protracted. Thus Fourier transform spectroscopy (FTS) did not become viable until mid-century when digital computers, which could readily become the required computations, were available. And, in the last decade, with the availability of inexpensive, dedicated microcomputers, Fourier transform spectrometers have proliferated.

The reason for FTS's popularity is its tremendous speed compared to the dispersive spectroscopy with a monochromator; in interferometry, data from all of the spectral frequencies are measured simultaneously, whereas the monochromator provides information only in a narrow band at a given time. This advantage in speed is designated as Fellgett's advantage. Other advantages of FTS may include convenience in experimentation, wider range of applicability, and high resolution.

Fourier transform spectroscopy results from the application of Fourier transform mathematics to the data of the interferogram. At the core is the basic equation of FTS, which links the spectrum to the interferogram through the Fourier transform. A detailed derivation of this equation will be given in due time.

This report is organized into three parts. The first, mathematical background (Chapters 1-9), presents a selective review of Fourier transform mathematics which should suffice for the reader to follow the subsequent derivations in FTS. No mathematical proofs are provided here; if desired, they may be found in mathematical texts. The second part (Chapters 10-12) treats the physics of the interferometry. The third part (Chapters 13-18) is on FTS itself, and it follows readily by applying part 1 to part 2.

Fourier transform spectrometers can vary in constructional and operational details, according to need. Notwithstanding, we will consider only the basic instrument used in the common mode. The variations are in the realm of esoterica.

In mathematical as well as physical treatment, recourse is sometimes made to idealized situations which are not exactly realizable. This greatly facilitates tractability, while having no serious effect on the results. For example, in deriving the basic equation of FTS an ideal beam splitter (50% reflectance, 50% transmission) is assumed and some factors are neglected. Notwithstanding, the deviations are easily correctable and do not call for revision of the basic equation, merely modification. In summary, the present treatment is simplified and basic, but valid and viable.

2. THE CONCEPT OF FREQUENCY

Frequency is the reciprocal of a direct measure in space or time. Thus, if the direct measure is space (cm), the reciprocal is designated spatial frequency (cm^{-1}); if the direct measure is time(s), the reciprocal is designated temporal frequency (s^{-1}).

Table 1 gives relevant examples of spatial and temporal frequencies, including notation and units.

TABLE 1: EXAMPLES OF SPATIAL AND TEMPORAL FREQUENCES

SPACE	SPATIAL FREQUENCY
Length (cm)	Length^{-1} (cm^{-1})
x = arbitrary length	$\xi = x^{-1}$
δ = optical path difference	$\xi = \delta^{-1}$
λ = optical wavelength	$\nu = \lambda^{-1}$ = wave number

TIME	TEMPORAL FREQUENCY
t = arbitrary time(s)	$f = t^{-1}$ (s^{-1}, cycles/s, Hz)

3. LINEAR SHIFT-INVARIANT SYSTEMS

A system may be defined as an assembly consisting of an input, a stimulus acting on the input, and an output. A system may be linear or shift-invariant; if both, the system is said to be linear shift-invariant (LSI).

Figure 1 illustrates the concept of linearity. Here $f_1(x)$ and $f_2(x)$ are two arbitary functions. The resultant function, or output, is just the sum of those two functions, i.e., $F(x) = f_1(x) + f_2(x)$.

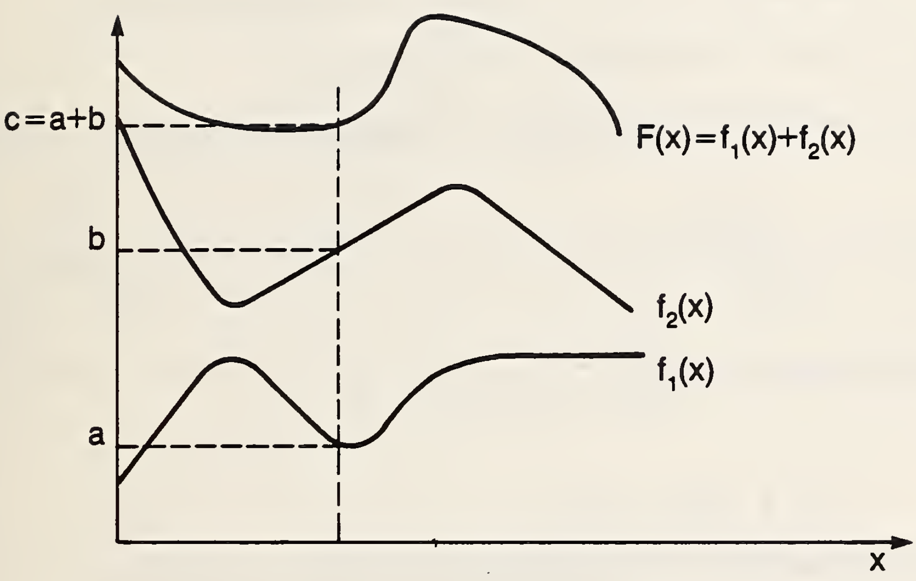


Figure 1. Diagrammatic example of a linear system.

Figure 2 illustrates the concept of shift invariance. The curve on the left represents an arbitrary function; that on the right the same function after being shifted. Note that the shape and height of the curve have not changed; in other words, the function is shift invariant. In the present work all systems are linear shift-invariant.

4. CONVOLUTION

Convolution as both a physical process and as a mathematical operation is of widespread importance, e.g., in measurement theory, statistics, and computing. Convolution is an integral part of Fourier transform mathematics, and in the applications of Fourier transforms it is ubiquitous.

The convolution of two functions f(x) and g(x) is given by the integral

$$\int_{-\infty}^{\infty} f(u)\, g(x-u)\, du = h(x), \tag{1}$$

where u is a dummy variable, and in symbolic notation by

$$f(x) * g(x) = h(x)\ . \tag{2}$$

There are four equivalent methods for convoluting: 1) direct integration, 2) numerical evaluation (sequences), 3) FFT (Fast Fourier Transform) of f(x) times FFT of g(x), then the inverse FFT, and 4) graphical construction. The last method will be treated here, after some preparatory remarks, as it provides insight into the process of convolution.

In English usage the term convolution has several allied meanings, of which 1) a form folded upon itself, and 2) the process of complicatedly moving from place to place can also describe the process of convolution in the

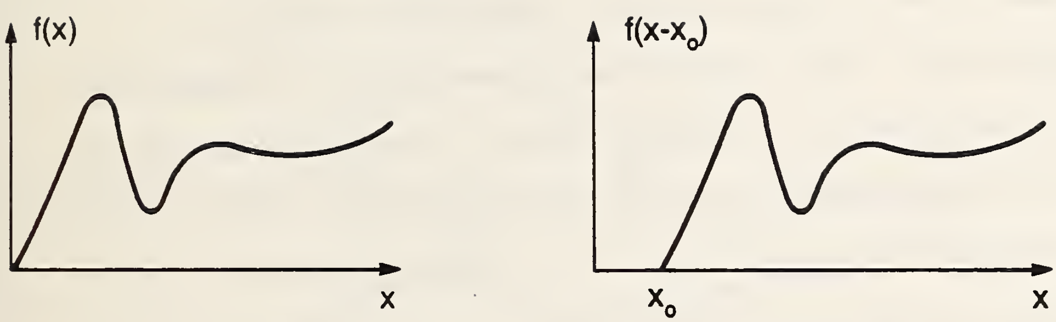


Figure 2. Diagrammatic example of a shift-invariant system.

physical-mathematical sense.

The process of convolution in effect is the scanning, or moving across, of one function by another function to yield a third, resultant function. In general, both the scanning function and the scanned function are distributions, i.e., some function of the same variable. Hence, in general, the resultant function is also a distribution (of the same variable); in fact it is a smoothed, spread-out version of the scanned function. Because we are dealing with LSI systems it will be found that it makes no difference which of the original, or convolving, functions is chosen to be the scanning function -- the final result will be the same.

Figure 3 illustrates the process of convolution of two arbitrary functions f(x) and g(x), shown as the two upper curves. Immediately below, the convolution integral $h(x) = f(x) * g(x)$ is shown represented by a shaded area; note that the function g(x-u) is a reflection of g(x). That a reflection can be conceived of as a fold, gives rise to the term Faltung (Ger. for fold) as an alternative designation for convolution. The lowermost curve is of h(x); the ordinate indicated by the solid vertical line is equal to the shaded area of the curve above. Note that the function h(x) is a smoothed and spread-out version of f(x); if g(x) had not been a smooth function, h(x) would have been less smooth.

We proceed now to carry out convolutions by graphical construction[+], and

---

[+]The beginner especially will find it convenient to draw one function backwards on a moveable strip of paper which can then be slid across the other function. Use of rectilinear coordinate paper and a light table are additional aids.

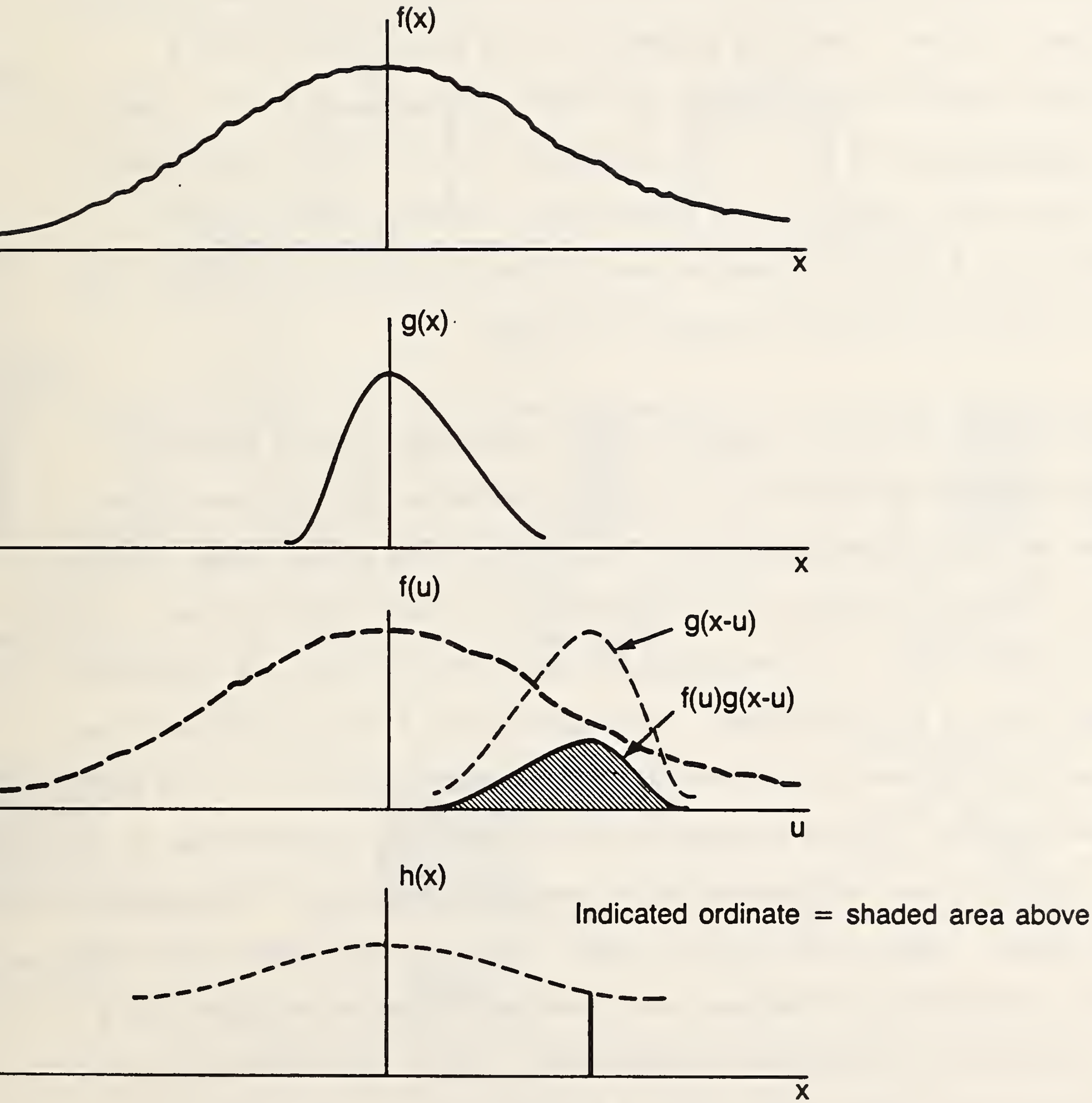


Figure 3. The convolution integral $h(x) = f(x) * g(x)$ represented by a shaded area (after Bracewell).

it will be evident that the method is just the visual manifestation of the operative convolution integral. The key to the method is to find the area of the product f(u)g(x-u) as x is allowed to vary.

Four examples, in order of increasing difficulty, are presented in Figures 4 through 7, below. The first (Figure 4) is the self-convolution of a rectangular-shaped function, which can assume values only of 0 or 1. Thus, in this special case, the area of the product of the functions f(u) and g(x-u) is the same as their common area of intersection. Further, owing to the symmetry, the function is invariant with reflection about the line x=0. The small arrow (↓) at x=0 will be used as an index.

The graphic convolution process begins on the second row of Figure 4, where the dashed form denotes the scanning function and the continuous form denotes the function being scanned. The abscissa value of the latter function which is opposite the index (↓) is the value of x which relates to the function h(x); i.e., it is the particular value of x for which the convolution integral h(x) is being evaluated. The index shows x = -1, and as there is no intersection, the area (A) is zero; therefore, h(x)= A; h(-1) = 0. The next diagram shows the scanning function shifted to the right such that now x = -.75 and the area of intersection is one-fourth the area of the function being scanned. Thus, h(x) = h(-.75) = .25. The scanning proceeds stepwise as shown in the sequence of diagrams, h(x) being determined at each step (x), until the area of intersection is again zero. Finally, to construct h(x) for all x-values, one simply plots h(x) = A as ordinate vs. x as abscissa, as shown in the last diagram. Here, the points h(-1) = 0; h(-.75) = .25; h(-.5) = .5, etc., obtained previously by graphical construction, are shown,

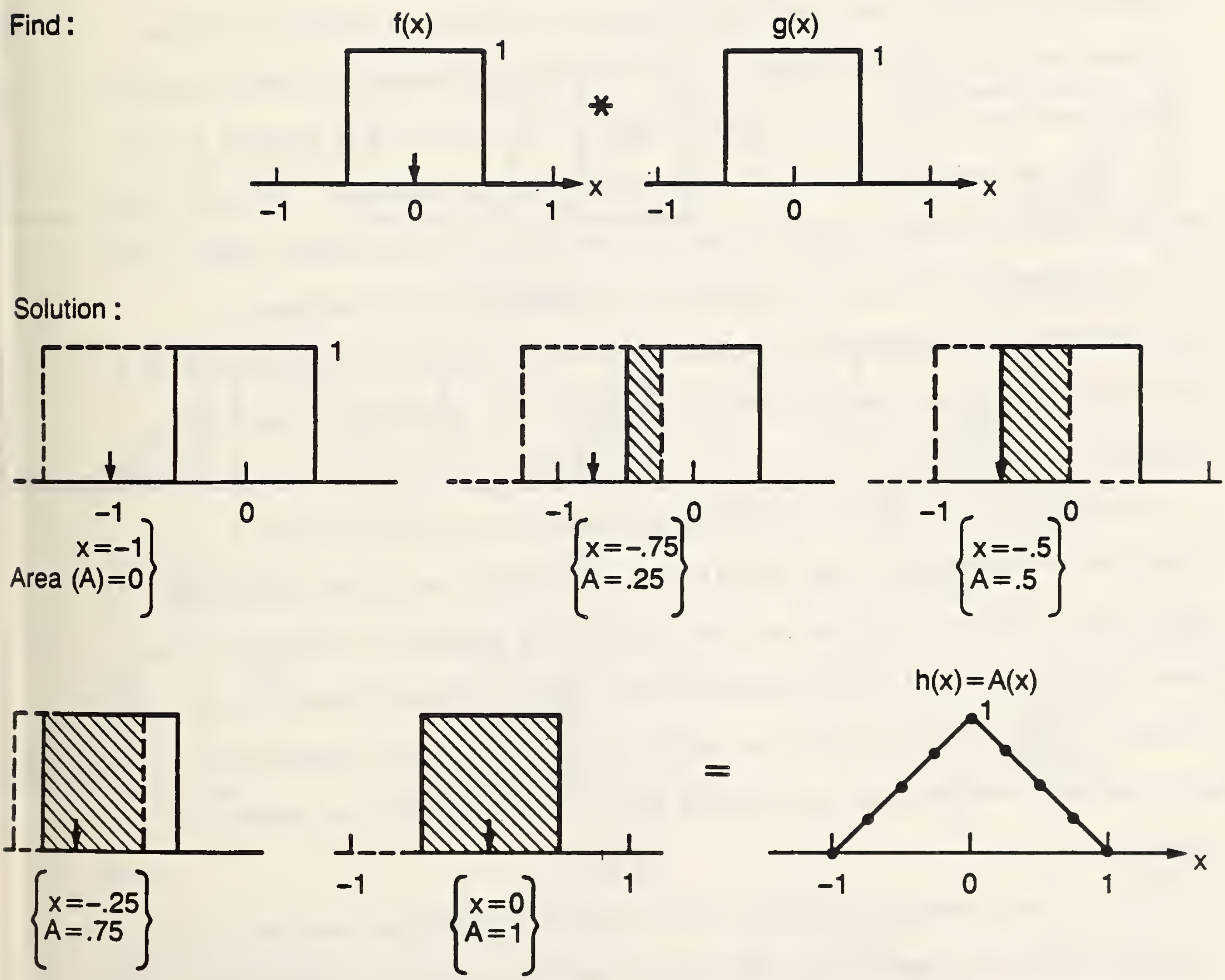


Figure 4. Convolution of functions by graphical construction, 1st example.

and the curve connecting these points is found to be triangular in shape. Note that this function, h(x), is, as expected, spread out compared to the original rectangular function.

The example of Figure 5 is slightly more difficult than the previous. Neither function to be convolved is symmetrical about the line x=0, so the convolution process starts with one function, arbitrarily chosen to be the rectangular one, being reflected. The succeeding procedure is the same as before, and again we have the special case of the area of the product of the functions and the common area of intersection being identical. The total area of the function being scanned is seen to be unity, so the various common areas of intersection can be easily computed, as indicated in the sequence of diagrams together with the corresponding x-value. A plot of these areas as a function of x yields the convolution integral h(x), as shown in the last diagram.

The example of Figure 6 considers the general case of convolution. Both functions are asymmetric, so the solution starts with one being reflected about the line x=0. Further, as indicated by the sequence of diagrams, the area of the product of the functions, which is the relevant area, shown shaded, is not identical with the intersection. The last diagram shows h(x) = f(x)*g(x), obtained by plotting A(x) vs. x, noted in the preceding sequence.

In the final example, Figure 7, the convolving functions are similar to those of Figure 6, except for orientation, and the graphical convolution procedure parallels that of the previous example. Note, however, that the shaded area, and hence, h(x), differ markedly in the two examples cited.

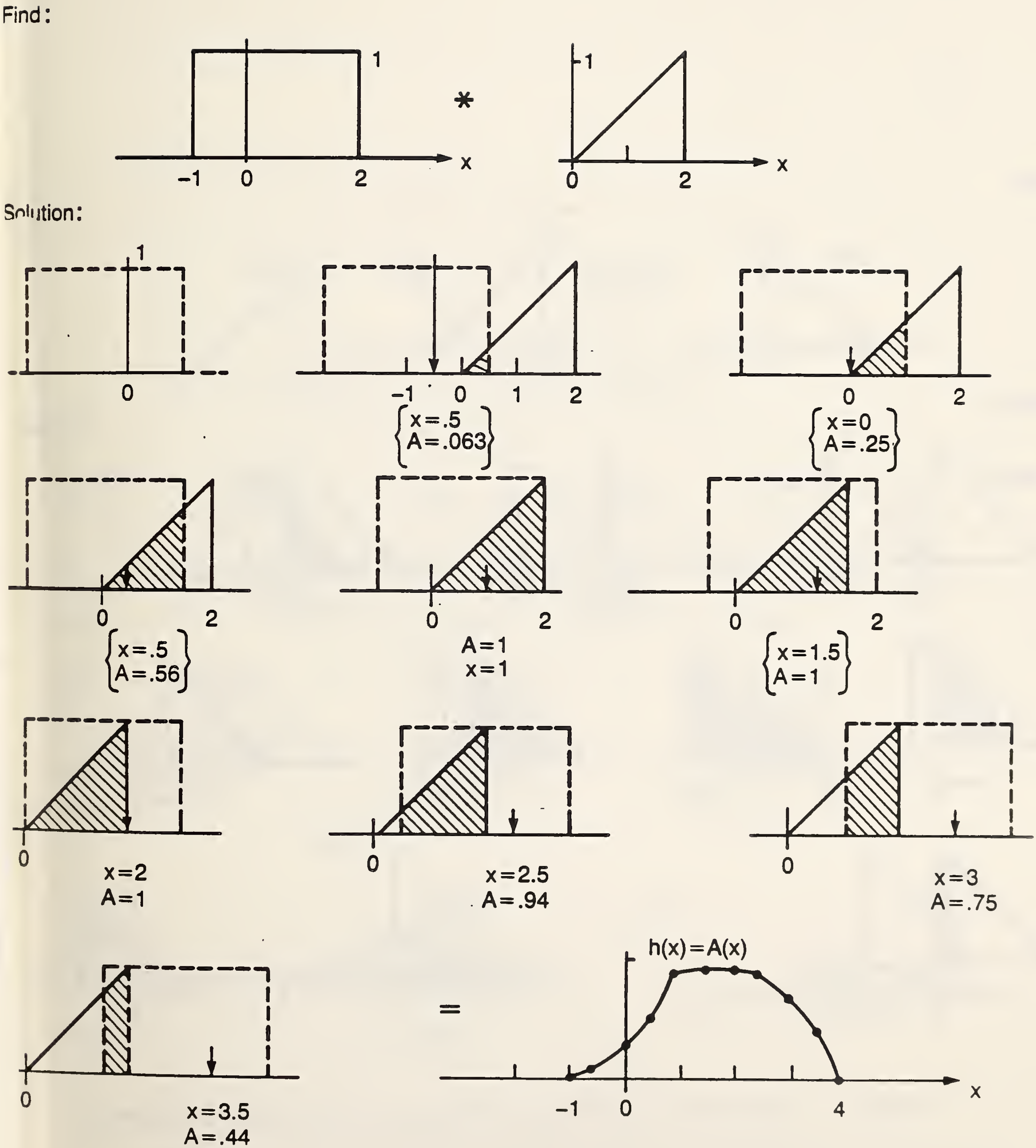


Figure 5. Convolution of functions by graphical construction, 2nd example.

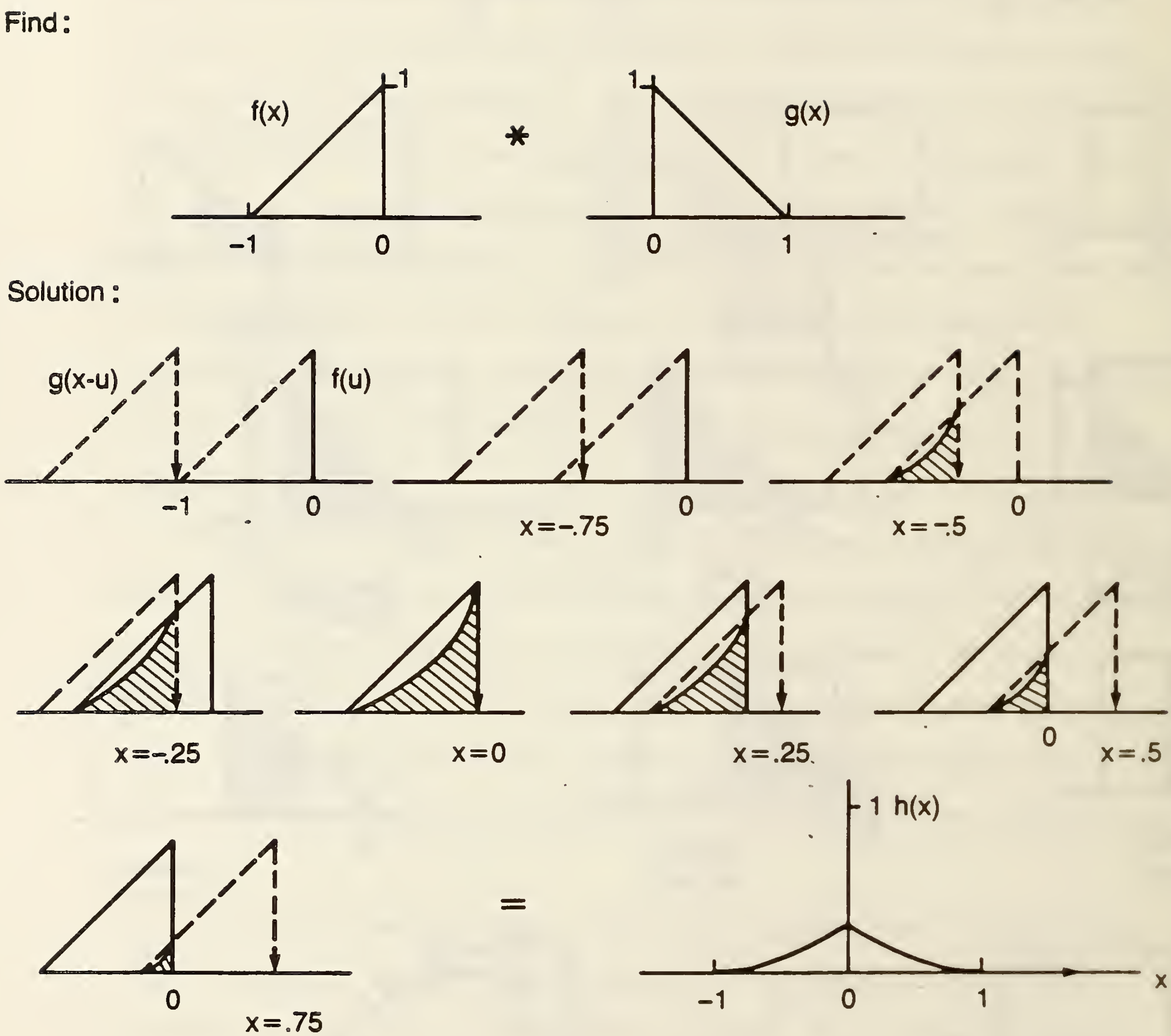


Figure 6. Convolution of functions by graphical construction, 3rd example.

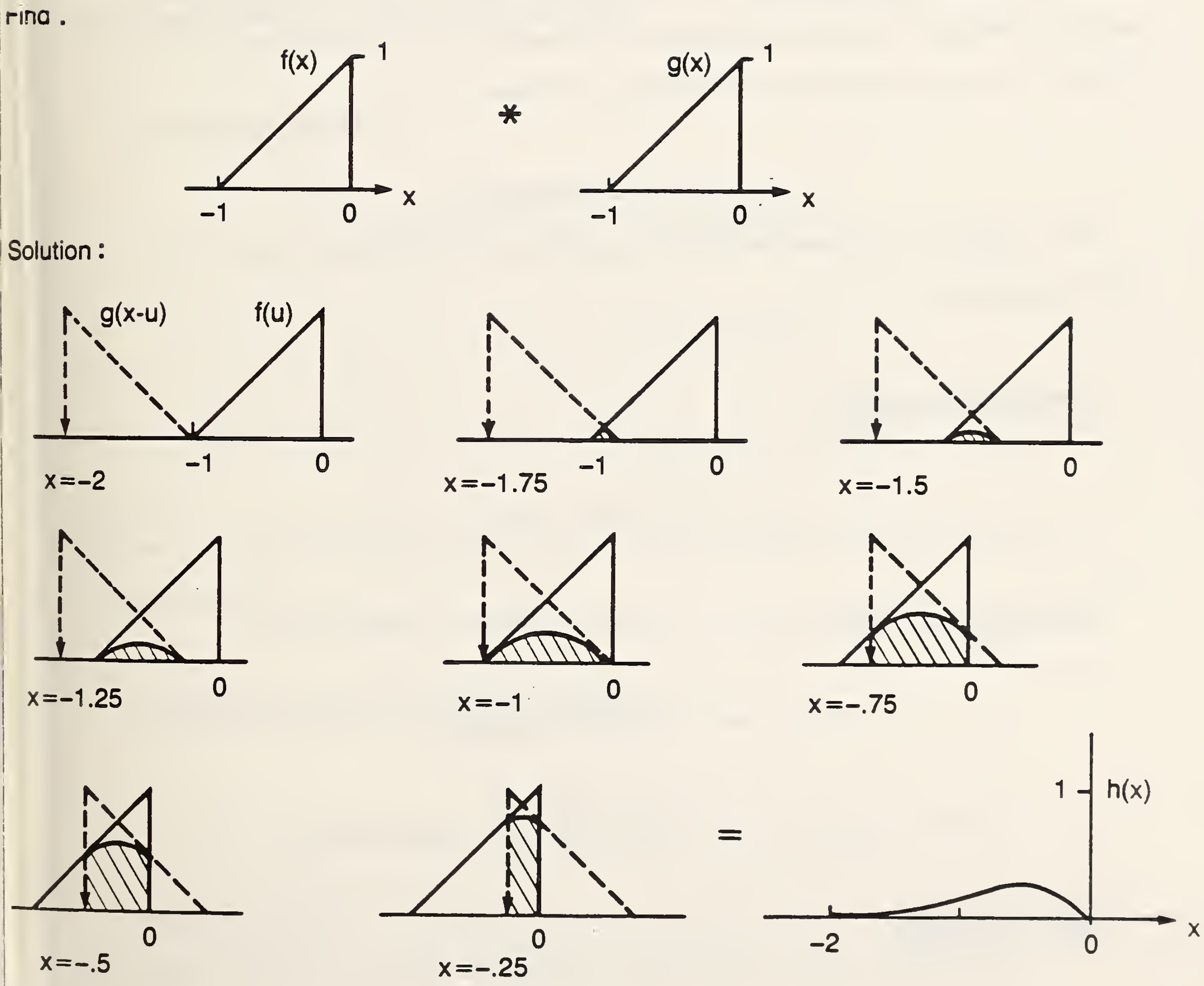


Figure 7. Convolution of functions by graphical construction, 4th example.

Problems for the diligent:

1. Denoting the rectangular-shaped functions of Figure 4 by rect(x), find rect(x) * rect(x) * rect(x) * rect(x). Does the result, h(x), resemble a Gaussian distribution?

2. The area under a convolution h(x) is equal to the product of the areas of the functions being convolved. This relationship provides a useful check. Do the examples of Figures 4-7 meet this criterion? Is the area under h(x) of Figure 6 the same as the area under h(x) of Figure 7?

5. PROPERTIES OF CONVOLUTION

This section lists some important properties of convolution. Most are self-explanatory.

<u>Commutative property</u>

$$f(x) * g(x) = g(x) * f(x)$$

<u>Distributive property</u> (a.k.a. linearity property of convolution)

With a and b arbitary constants,

$$[af(x) + bg(x)] * h(x) = a[f(x)*h(x)] + b[g(x)*h(x)]$$

<u>Shift invariance</u>

Given $\quad f(x) * g(x) = h(x),$

then $$f(x-x_0) * g(x) = h(x-x_0)$$

Similarly, $$f(x) * g(x-x_0) = h(x-x_0)$$

In other words, if either function, $f(x)$ or $g(x)$, is shifted by an amount x_0, the resulting convolution is simply shifted by the same amount, while remaining unchanged in magnitude and form.

Associative property

$$[f(x)*g(x)] * h(x) = f(x)*g(x)*h(x)$$

$$= f(x)*h(x)*g(x)$$

$$= h(x)*f(x)*g(x), \text{ etc.}$$

In other words, the order in which individual convolutions are performed is inconsequential.

Convolution property of delta functions

$$f(x)*\delta(x) = f(x)$$

The convolution of a δ-function with any other function merely reproduces the other function; this may be used as a definition of the δ-function.

6. DECONVOLUTION

It will become evident later (Part 3) that convolutions are an ineluctable characteristic of Fourier transform spectrometry which may render the resulting data, or output, useless. Thus, the need will arise to undo convolutions, i.e., to deconvolute.

The concept of deconvolution is a simple one. Suppose an LSI system of the form

$$f*g = h,$$

where the functions f, g, and h represent input, impulse response, and output. Given g and h, the problem is to determine f.

Solution: Let g^{-1} be the reciprocal of g, i.e.,

$$g*g^{-1} \equiv 1 \tag{3}$$

Then,

$$f*g*g^{-1} = h*g^{-1}$$

$$f = h*g^{-1}.$$

Similary, if

$$f^{-1}*f \equiv 1$$

$$g = h*f^{-1} \; .$$

Thus, given a convolution product of two functions, or output function, one of the convolving functions can be approximately determined by deconvolution of the output function with the other convolving function.*

Operationally, deconvolution is a divisional process, while convolution is multiplicative. And while convolution can be done graphically, and sometimes the process even visualized, deconvolutions are more complicated and best left to be caried out by a computer subprogram+.

7. FOURIER TRANSFORMS

Fourier transforms are obviously essential to the conduct of Fourier transform spectroscopy, and that alone would justify its importance. But Fourier transforms are vital in other pursuits as well; e.g., electrical signal analysis, diffraction, optical testing, optical processing, imaging, holography, and remote sensing. Thus, a knowledge of Fourier transforms can be a springboard to many allied fields.

The idea behind Fourier transforms is that a function of direct space (or time) can be expressed equivalently as a complex-valued function of reciprocal space, i.e., frequency (sometimes called Fourier space.) Generally, the

---

*Deconvolution amplifies noise, resulting in unphysical results. (e.g., peaks narrower than the instrumental resolution.) Some form of smoothing is always required, and selection of the appropriate smoothing is not straightforward.

+Actually, deconvolution can be caried out manually by long division or with the aid of a hand calculator (see Bracewell, cited in Bibliographiy). Notwithstanding, these methods are only of possible academic interest.

ensuing mathematical operations greatly simplify solutions to problems, even those otherwise recondite.

Fourier transforms may be defined by the following two integrals:

$$f(x) = \int_{-\infty}^{\infty} F(\xi) e^{j2\pi x\xi} \, d\xi \; ; \tag{4}$$

$$F(\xi) = \int_{-\infty}^{\infty} f(x) e^{-j2\pi x\xi} \, dx \; . \tag{5}$$

$F(\xi)$ is said to be the Fourier transform of $f(x)$, and $f(x)$ is said to be the <u>inverse</u> Fourier transform of $F(\xi)$. $F(\xi)$ and $f(x)$ are said to comprise a Fourier transform pair. Note that both integrals are similar, but differ in the polarity of $j = \sqrt{-1}$.

In symbolic notation $\mathcal{F}$ stands for "the Fourier transform of ...". Thus, for example, $\mathcal{F}\{F(x)\} = F(\xi)$ means the Fourier transform of $f(x)$ is $F(\xi)$. Similarly, $\mathcal{F}^{-1}$ stands for "the inverse Fourier transform of ...", and $\mathcal{F}^{-1}\{F(\xi)\} = f(x)$ means the inverse Fourier transform of $F(\xi)$ is $f(x)$. It is easy to show from the definitions of $f(x)$ and $F(\xi)$ that

$$\mathcal{F}^{-1}\mathcal{F}\{f(x)\} = f(x) \quad , \tag{6}$$

and

$$\mathcal{F}\mathcal{F}^{-1}\{F(\xi)\} = F(\xi) \; . \tag{7}$$

Thus, the inverse Fourier transform of the Fourier transform of a function is just the function itself; in other words, $\mathcal{F}^{-1}\mathcal{F}$ and $\mathcal{F}\mathcal{F}^{-1}$ can be considered as products equal to unity.

In the (Fourier) integral of equation (4), $F(\xi)$ serves as a complex-valued weighting factor prescribing the correct amplitude and phase for all of the exponential components of which $f(x)$ is comprised. In equation (5) the converse relationship exists.

On the other hand, if the function $f(x)$ is real and even $[f(x) = f(-x)]$ -- a condition met in conventional FTS -- by invoking Euler's formula+, equation (5) can be simplified to

$$F(\xi) = \quad f(x)\cos(2\pi\xi x)\ dx \quad . \tag{8}$$

This expression may be interpreted as follows: at any particular frequency $\xi=\xi_1$, the value of $F(\xi_1)$ is equal to the area of the product $f(x)\cos(2\pi\xi_1 x)$, and $F(\xi) = F(\xi_1)+F(\xi_2)+F(\xi_3)+\ldots\ldots\ldots$

By similar argument, equation (4) reduces to

$$f(x)= \quad F(\xi)\cos(2\pi\xi x)d\xi \ . \tag{9}$$

Equations (8) and (9) are known as the cosine-form of the Fourier transform equation, or of the Fourier integral.

There are a great many Fourier transform pairs, and each is unique. However, only a few simple, commonplace functions are cogent here and need to be considered. Table 2 lists the functions by name and notion. Figure 8 is a pictorial collection of these Fourier transform pairs shown side-by-side. Owing to the reciprocity of Fourier transform pairs it would be redundant to

---

+See e.g., I. S. Sokolnikoff and E. S. Sokolinkoff, Higher Mathematics for Engineers and Physicists, McGraw-Hill, New York, 1941.

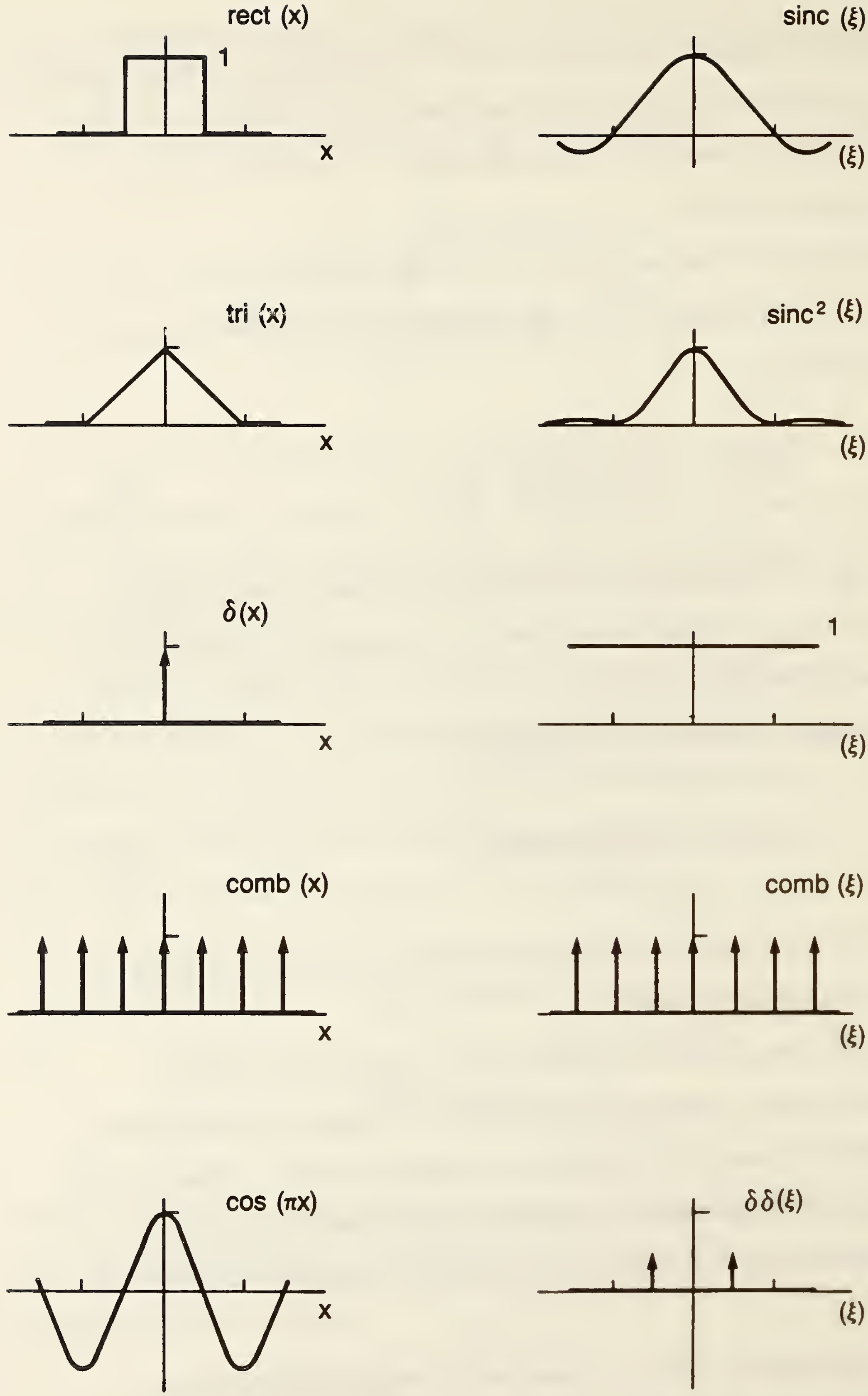


Figure 8. Some Fourier transform pairs. The ticks show where the variables have a value of unity. Impulses are denoted by arrows of a length equal to the strength of the arrow.

repeat the drawings; e.g., if $\mathcal{F}$[rect(x)] = sinc(ξ), then $\mathcal{F}$[sinc(x)] = rect(ξ), etc.

TABLE 2. COMMON FUNCTIONS

FUNCTION	NOTATION
Rectangle/boxcar	$\text{rect}(x) = \begin{cases} 1 & \lvert x\rvert < 1/2 \\ (1/2 & \lvert x\rvert = 1/2) \\ 0 & \lvert x\rvert > 1/2 \end{cases}$
Triangle	$\text{tri}(x) = \begin{cases} 1 - \lvert x\rvert & \lvert x\rvert < 1 \\ 0 & \lvert x\rvert > 1 \end{cases}$
Delta function/impulse symbol	δ(x)⁺
Comb	comb(x) = δ(x-n), n = 0,1,2,
Even impulse pair	$\delta\delta(x) = (1/2)\delta(x+1/2) + 1/2\ \delta(x-1/2)$
Sinc	$\text{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$⁺⁺

⁺See properties of the delta function, infra.

⁺⁺This definition of the sinc function, due to Bracewell, has been adopted here because of its convenience. An earler definition still used by some other authors is sin x/x.

8. PROPERTIES OF FOURIER TRANSFORMS

This section lists some properties of Fourier transforms. Particular attention is directed to the transform of a convolution and its conjugate, the transform of a product, perhaps the most significant of all Fourier transform properties. As before, $\mathcal{F}\{f(x)\} = F(\xi)$; $\mathcal{F}\{g(x)\} = G(\xi)$.

Linearity property

The transform of a sum of two functions is the sum of their individual transforms.

$$\mathcal{F}\{f(x)+g(x)\}=F(\xi)+G(\xi).$$

Thus in LSI systems the spectrum of a sum of signals can be computed by merely adding together their individual spectra.

Transform of a transform

If

$$\mathcal{F}\{f(\xi)\} = F(x),$$

then

$$\mathcal{F}\{F(x)\} = f(-\xi).$$

Transform of a convolution

The Fourier transform of a convolution is given by the product of their individual transforms.

$$\mathcal{F}\{f(x) * g(x)\} = F(\xi)G(\xi).$$

Transform of product

The Fourier Transform of a product is given by the convolution of their individual transforms.

$$\mathcal{F}\{f(x)g(x)\} = F(\xi)*G(\xi).$$

Properties of the delta function

The δ-function may be described qualitatively by a pulse whose amplitude tends to infinity as its width tends to zero, and the area of the function is constrained to be unity. Hence, it is not possible to graph the actual function (although it can be pictorially shown as a sequence of pulses). Nevertheless, the δ-function can be symbolically represented by a pictograph when required for convolution by graphical construction. A property of the δ-function is that when it is convolved with any other function, it merely replicates that other function. Thus, the pictograph of the δ-function, is an impulse of unit strength.

<u>Properties of the comb function</u>

The comb function is so useful it merits special attention. This function may be considered as an extension of the δ-function; indeed the comb function is just a linear array of δ-functions, and it is used extensively to perform either of two important operations: 1) sampling, or 2) replication.

The product of a comb function with another function samples the other function; the sampling property of the comb function is illustrated in Fig. 9.

The convolution of a comb function with another function replicates an infinite array of the other function (the comb function is of infinite extent); the replicating property of the comb function is illustrated in Fig. 10.

Problem:

$$\text{Find } \mathrm{sinc}(x) * \mathrm{sinc}(x)$$

9. THE FOURIER TRANSFORM AND LINEAR SHIFT-INVARIANT SYSTEMS

Earlier, this report defined a system as an assembly consisting of an input, a stimulus acting on the input, and an output; and the concept of the

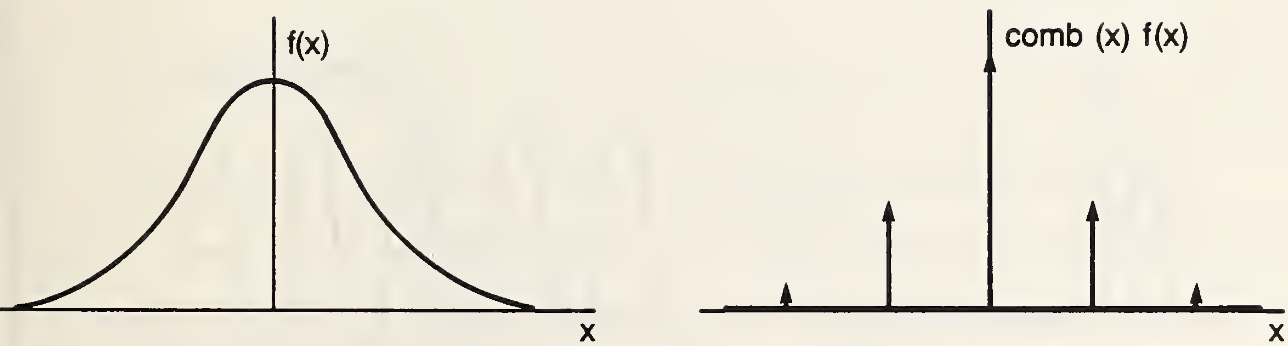


Figure 9. The sampling property of comb (x).

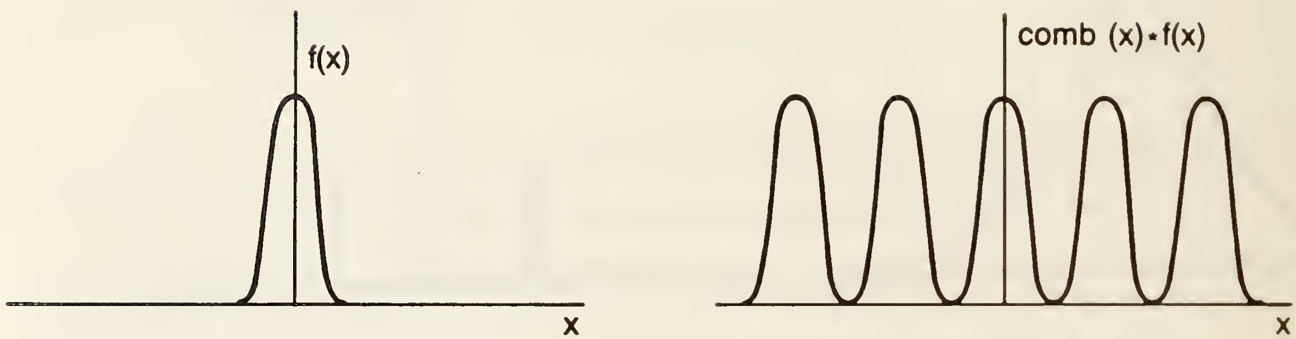


Figure 10. The replicating property of comb (x).

linear shift-invariant system, in particular, was introduced. Now the significance of LSI systems can be explored in the light of Fourier transforms. First, it should be mentioned that a system need not be linear shift invariant for Fourier transforms to be applicable, however, if the system is linear shift invariant, as required here, the mathematics is greatly simplified.

Assume an LSI system having an input f(x), a stimulus g(x), and an output h(x); Fig. 11 is a schematic representation of the LSI system. Now, it is a property of LSI systems that given an input and an output, a function can be found which when convoluted with the input will yield the output; i.e.,

$$f(x) * g(x) = h(x) \quad . \tag{10}$$

The function relating the input and the output is the stimulus g(x), better known as the impulse response. The simplicity of eq. (10) belies its potency; e.g., if the impulse response of an LSI system is known, then the output of the system can be determined for a variety of inputs. Later it will be shown that given any two of the three functions defining the LSI system, it is possible to determine the remaining function.

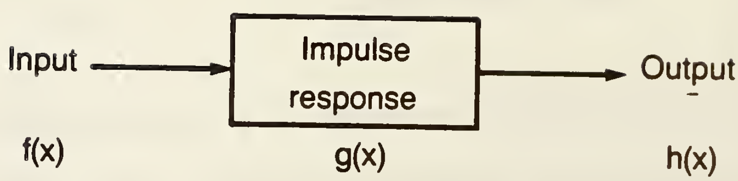


Figure 11. Schematic representation of a linear shift-invariant system.

An alternative, and generally more useable, form of the above equation can easily be derived therefrom:

With input f(x), impulse response g(x), and output h(x), let

$$\mathcal{F}\{f(x)\} = F(\xi),$$

$$\mathcal{F}\{g(x)\} = G(\xi)\ ,$$

and

$$\mathcal{F}\{h(x)\} = H(\xi)\ ,$$

where $F(\xi)$ is the spectrum of the input, $G(\xi)$ is the spectrum of the impulse response, referred to as the <u>transfer function</u>, and $H(\xi)$ is the spectrum of the output.

Taking the Fourier transform of both sides of eq. (10),

$$\mathcal{F}\{f*x)*g(x)\} = \mathcal{F}\{h(x)\}\ ,$$

and recalling the transform properties of a convolution, viz.,

$$\mathcal{F}\{f(x)*g(x)\} = F(\xi)G(\xi)$$

one obtains

$$H(\xi) = F(\xi)G(\xi)\ . \qquad (11)$$

Thus, the output spectrum of an LSI system is given simply by the product of the input spectrum and the transfer function.

Further,

$$\mathcal{F}^{-1}\mathcal{F}\{h(x)\} = \mathcal{F}^{-1}\{H(\xi)\} \quad ;$$

$$h(x) = \mathcal{F}^{-1}\{H(\xi)\} \quad .$$

Thus, the output itself can be determined by the inverse Fourier transform of the output spectrum.

10. OVERVIEW OF THE FOURIER TRANSFORM SPECTROPHOTOMETER

The Fourier transform spectrophotometer is a sophisticated, complex system in the sense that it is comprised of several distinct components, each performing its own functions, but interrelated and working toward the common objective. Figure 12 is a block diagram of the Fourier transform spectrophotometer; an infrared light source shall be assumed because infrared is the usual region of application.

The interferometer, which is computer controlled, regulates the optical path difference between two beams of light causing interference in such manner as to produce a combined light beam of periodically-modulated intensity. (A detailed review of the Michelson interferometer is given in the next section.) The modulated light is detected by a sensor, a pyroelectric or

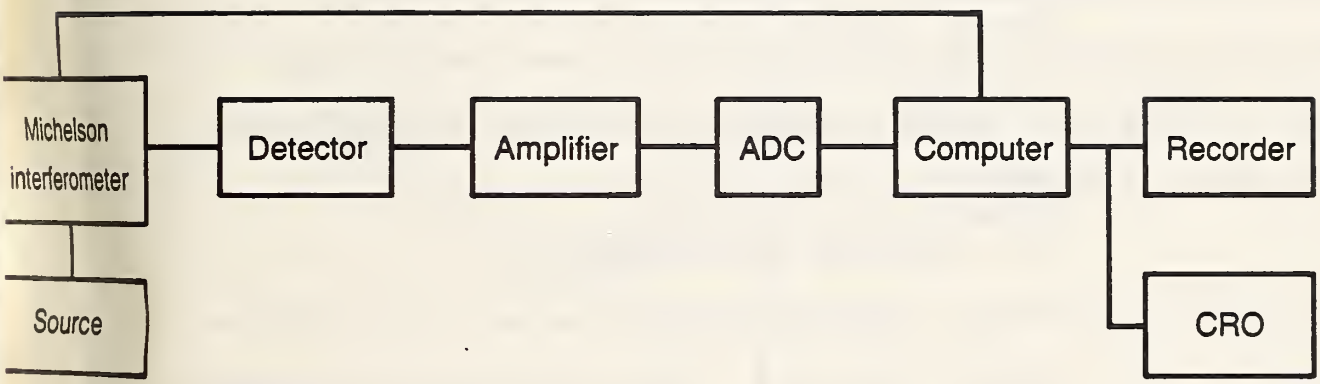


Figure 12. Block diagram of Fourier transform spectrophotometer.

an infrared photon detector, which converts the light to a (feeble) electrical signal which is then amplified.

The raw data are in the form of an interferogram which requires digitization before it may be transformed into a spectrum. Hence, the analog signal data are sampled and a computer-controlled analog-to-digital converter (ADC) digitizes and relays the digitized data to the computer for storage in memory. The computer performs a Fourier transform and plots the spectrum on a plotter and/or displays it on an oscilloscope. By comparing spectra obtained with and without a sample in the (combined) light beam, a function also performed by the computer, the spectrum of the sample itself is obtained.

Some errors may be introduced in the sampling process, owing to discretization, imperfect band-limiting of the data due to noise, and aliasing; see Chapt. 16. Further, phase corrections must be made for the nonlinear phase shifts which occur in an actual interferometer.

11. MICHELSON INTERFEROMETER

The "heart" of the Fourier transform spectrophotometer is the Michelson interferometer, and a schematic diagram of a simple Michelson interferometer is shown in Fig. 13. The apparatus consists essentiallly of an external light source, a beam splitter, and two mutually perpendicular plane mirrors, one fixed and one movable.

The interferometer is idealized by the following assumption: 1) the external source of radiation is a point source of monochromatic light; 2) the beam splitter transmits 50% of the light incident on it and reflects 50% of the light incident on it; 3) the plane mirrors are perfect (total) reflectors and are exactly mutually orthogonal.

In operation, a beam of radiation from the external source falls on the

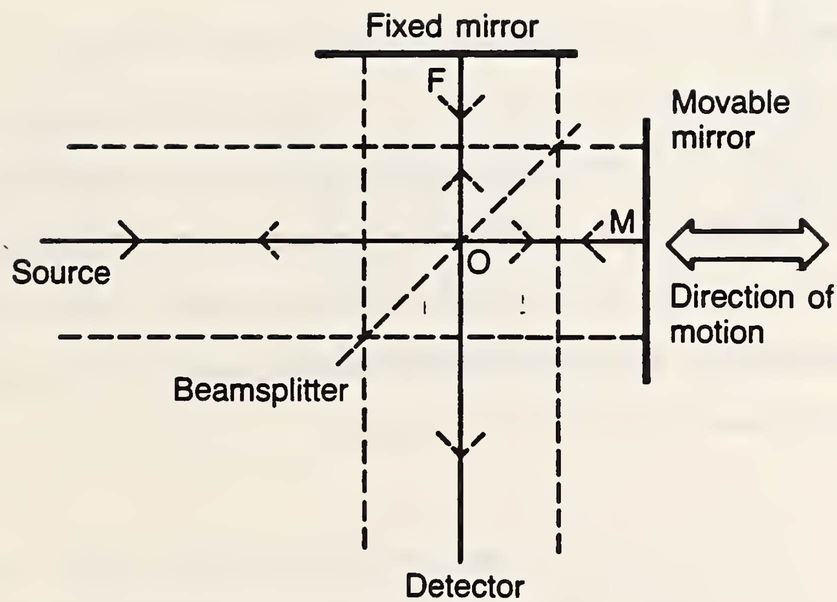


Figure 13. Schematic diagram of a Michelson interferometer (after Griffith).

beamsplitter which reflects half of the light to the fixed mirror (path OR) and transmits half of the light to the movable mirror (path OM). Each beam is then reflected, retracing its path back to the beamsplitter whence each is again partially reflected and partially transmitted. Thus, in general, a fraction of the light from the source is returned to the source, while the remaining fraction reaches the detector (D). Usually only the latter beam is of concern; note that the latter beam is comprised of two beams which have been recombined. The optical path difference between these two beams, δ--also called the retardation--is 2(OM-OF).

Figure 14 shows the effects of varying optical path difference. When OM=OF, $\delta=0$. Then both beams are perfectly in phase, and on recombination at the beam splitter, (total) constructive interference takes place; hence the resultant intensity of the light reaching the detector is quadruple the intensity associated with only a single one of these beams. All of the energy from the source reaches the detector and none is returned to the source. The condition $\delta=0$ is referred to as the null or zero.

When OM=OF+$\lambda/4$, $\delta=\lambda/2$, where λ is the wavelength of the light. Then both beams are exactly out of phase, and on recombination at the beam splitter, total destructive interference takes place; hence the resultant intensity of the light reaching the detector is zero. All of the energy from the source is returned to the source.

When $\delta=\lambda$, the waves are indistinguishable from those depicted for $\delta=0$. Similarly, when $\delta=3\lambda/2$, the waves are indistinguishable from those depicted for $\delta=\lambda/2$. Thus, the criterion for maximum intensity is given by $\delta=n\lambda$, where n is an integer, and the criterion for minimum (zero) intensity is given by $\delta=(n+1/2)\lambda$. For intermediate values of δ, correspondingly intermediate values

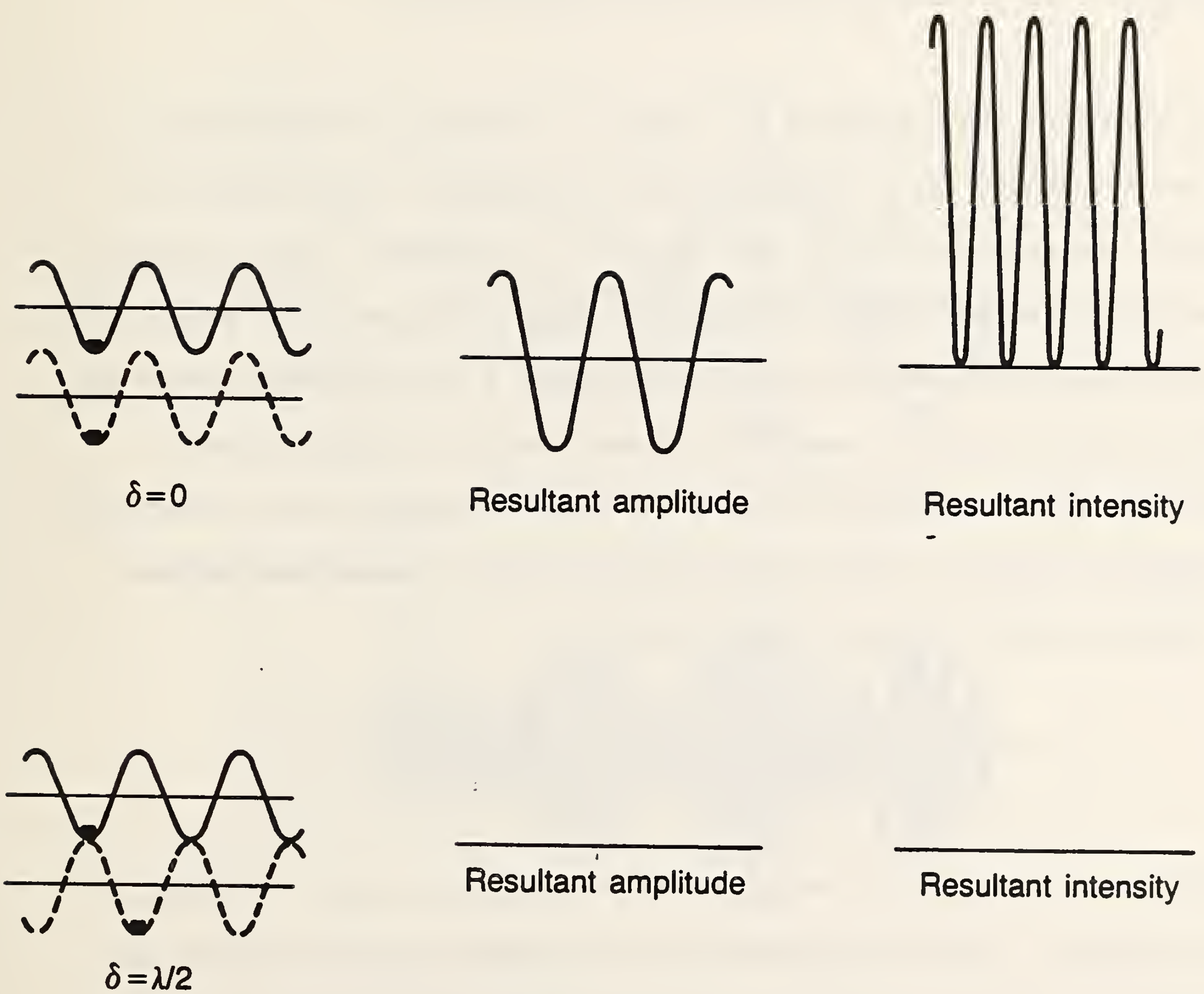


Figure 14. Effects of interference in a Michelson interferometer. The beam traveling to the fixed mirror is depicted by the solid line; the beam traveling to the movable mirror is depicted by the broken line; the marker denotes light which left the source at the same time: δ is the retardation. The resultant amplitude is the sum of the individual amplitudes and the resultant intensity is the square of the resultant amplitude.

of intensity will be obtained.

Now, consider the mirror to be moved from the null at constant velocity, while maintaining orthogonality with the fixed mirror. The retardation will be related to the velocity of the mirror, v, and the duration of its movement, t, by the relation $\delta = 2vt$. Thus, the movable mirror will cause periodic alternations of constructive and destructive interference.

The recombined light beam can be viewed as a function of the duration of the mirror movement or as a function of the retardation. When duration is considered, the moving mirror is seen to act as a sinusoidal signal generator which modulates the amplitude of the carrier wave. This amplitude modulation, or AM, is analogous to the familiar AM radio where a low-frequency modulating electrical signal is used to modulate the amplitude of a high-frequency carrier wave, and is depicted in Fig. 15. Figure 16 compares the intensity of the recombined beam as a function of retardation with the intensity of the same recombined beam as a function of time.

12. COHERENCE AND INTERFERENCE

Interference can be explained in either of two ways, phase difference or coherence theory. They are complementary and together form the subject of this chapter.

Light, like other forms of radiant energy, is electromagnetic. For most purposes analysis is made in terms of harmonic plane waves and solutions derived from them by superposition (addition). In addition to

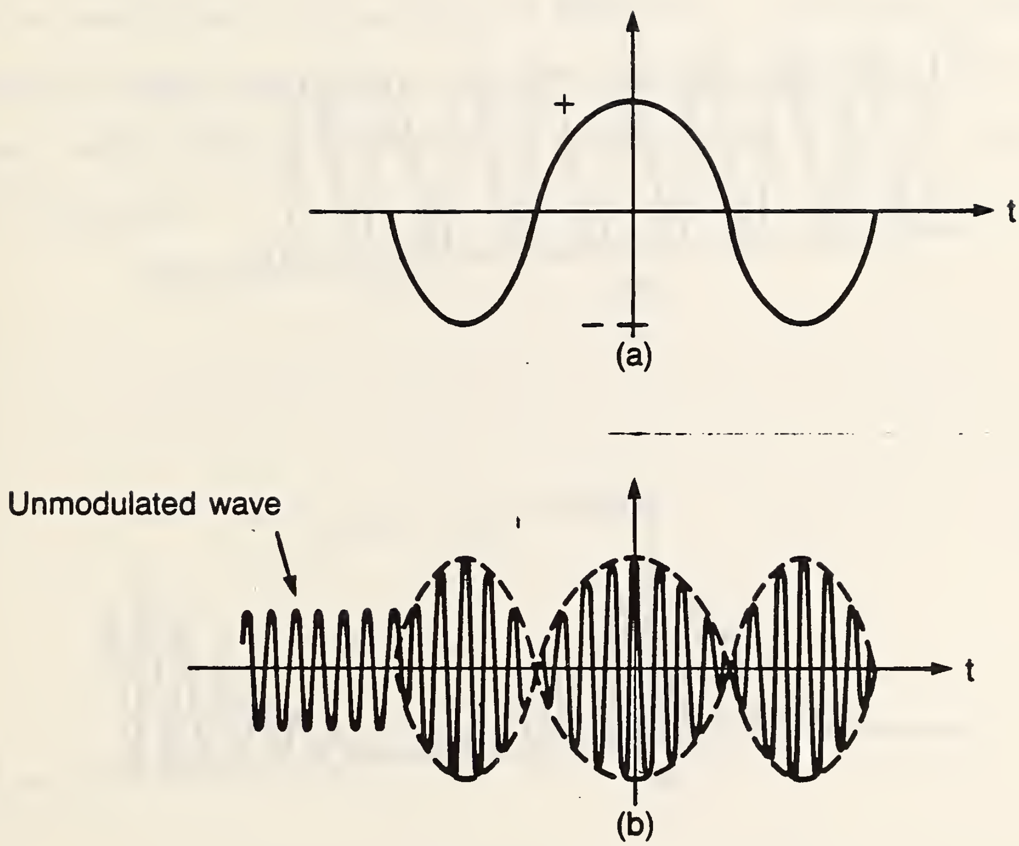


Figure 15. Amplitude modulation. (a) modulating signal: (b) modulated carrier wave.

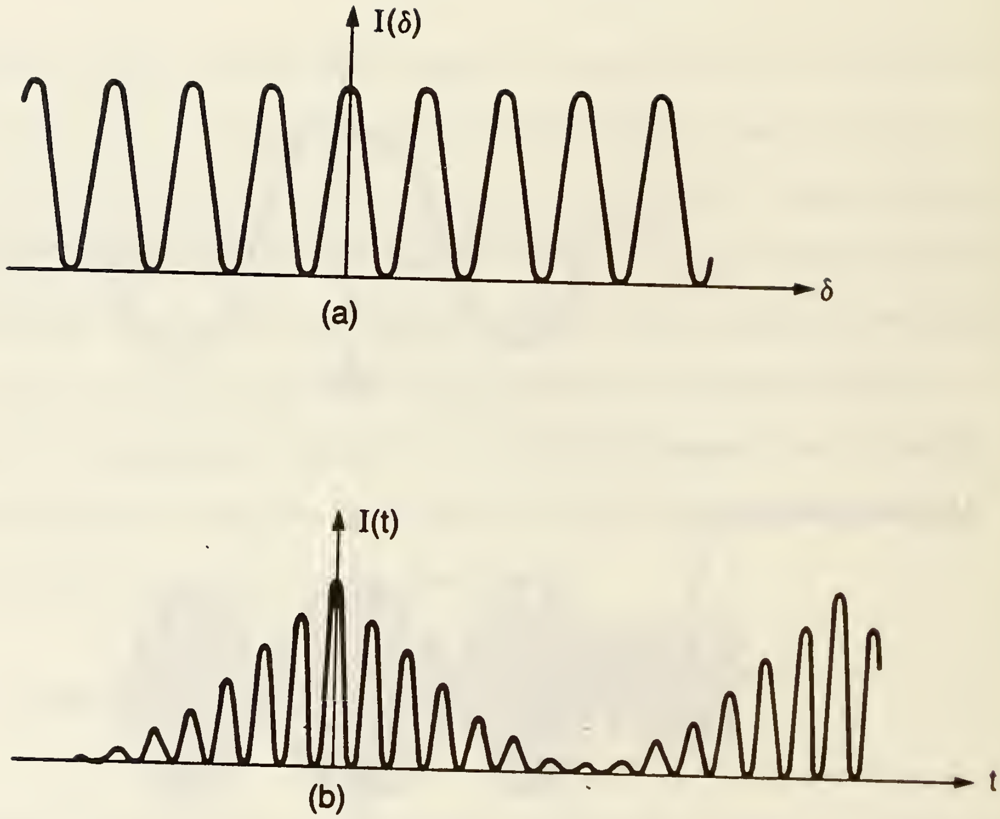


Figure 16. Two views of the same interference pheonmenon. (a) intensity as function of optical path difference (retardation); (b) intensity as function of time. The markers on each curve indicate exact counterparts.

wavelength, plane waves may be specified as to direction of propagation of the wave, the amplitude, and the phase, as shown in Fig. 17. Here the wave is assumed to be propagating in the z-direction, and the component of the displacement in a certain transverse direction, say, the x-direction, is the real part of

$$\vec{u} = a\ e^{j(-2\pi\nu z+\emptyset)} \quad , \tag{12}$$

which may be recognized as the equation of a sinusoid. The complex quantity $\vec{u}$ is called the vector amplitude, a is called the scalar amplitude, and $\emptyset$, the angle that $\vec{u}$ makes with the x-direction is called the phase angle, or more commonly, the phase.

Hence,

$$f(z) = a\ e^{j(-2\pi\nu z+\emptyset)} \quad , \tag{13}$$

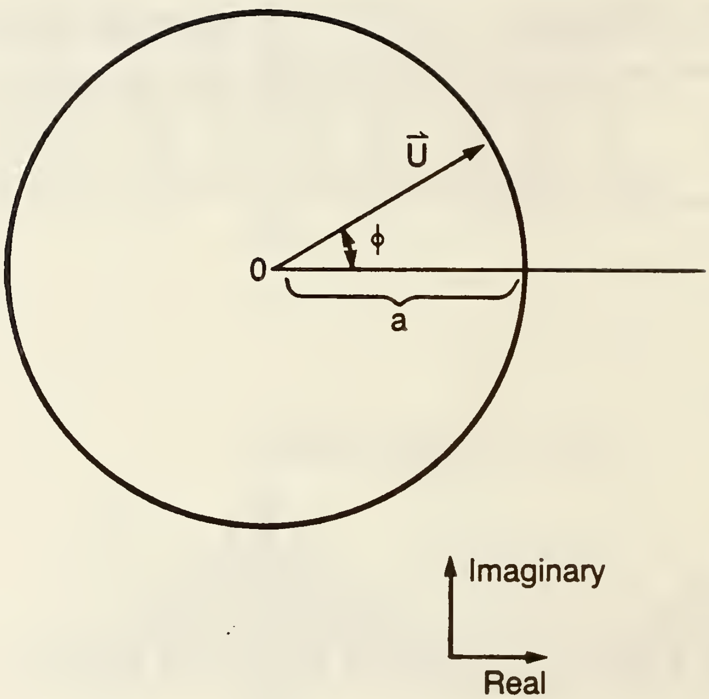


Figure 17. Vector amplitude diagram of one beam (after Burnett, et al).

where a is a real-valued function. A graph of the function f(z), given in Fig. 18, shows another interpretation of phase: it determines the position of the function projected along the direction of propagation.+ Note that the magnitude of the shift z_0 is related to ϕ by the equation

$$z_0 = \frac{\phi\lambda}{2\pi} , \tag{14}$$

or

$$\phi = 2\pi\nu z_0 \tag{15}$$

Substituting in the above equation for f(z) gives

$$f(z) = a\ e^{j2\pi\nu(z-z_0)} \tag{16}$$

The phase difference between two (or more) beams determines whether or not they will interfere. If the phase difference is time invariant, the beams will interfere and the light is said to be coherent. Otherwise the light would be incoherent or partially coherent.

Now consider two waves or beams traveling in the same direction, as shown in the vector amplitude diagram of Fig. 19.

+ If the relationship between Figs. 17 and 18 is confusing, try thinking of the vector $\vec{u}$ (Fig. 17) as rotating with constant angular velocity as it moves in the z-direction with constant linear velocity.

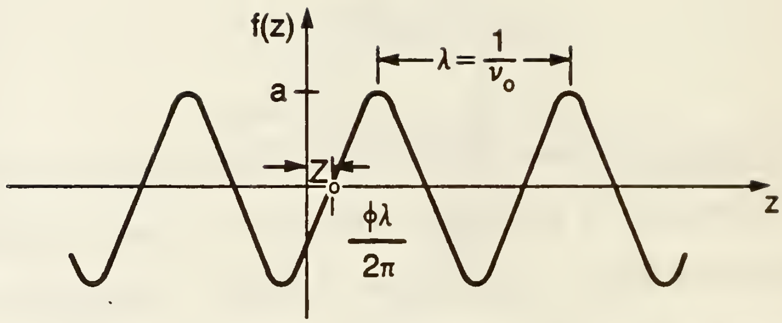


Figure 18. Sinusoidal function of amplitude a, frequency ν_0 and phase ϕ.

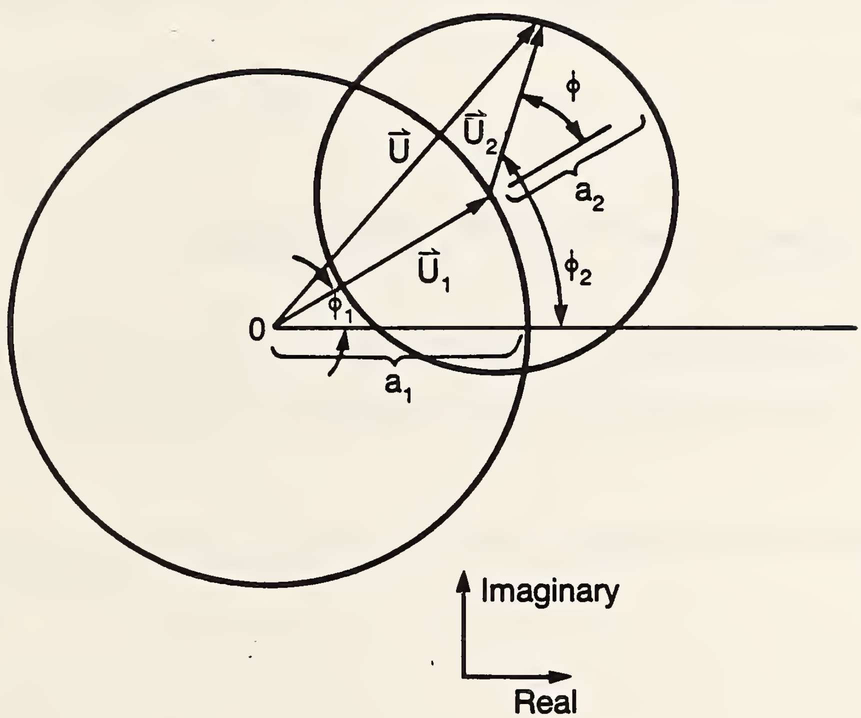


Figure 19. Vector amplitude diagram of two-beam interference (after Burnett, et al).

Here, $\vec{u}_1$, and $\vec{u}_2$ are the vector amplitudes of the two waves of constant phase difference $\phi = \phi_2 - \phi_1$, and $\vec{u}$ is the vector amplitude of the resultant wave.

The intensity of light is given by the product of the vector amplitude and the complex conjugate+ of the vector amplitude; i.e.,

$$I = \vec{u}\vec{u}^* \quad . \tag{17}$$

For two-beam interference as depicted in Fig. 19, the intensities of the beams considered separately would be $I_1 = \vec{u}_1\vec{u}_1^* = a_1^2$ and $I_2 = \vec{u}_2\vec{u}_2^* = a_2^2$,

---

+The complex conjugate of $ae^{j\alpha}$ is $ae^{-j\alpha}$, and vice versa.

and the intensity of the resultant would be

$$I_{2c} = (\vec{u}_1+\vec{u}_2)(\vec{u}_1*+\vec{u}_2*) \tag{18}$$

$$= I_1+I_2+2I_1^{1/2}I_2^{1/2}\cos\phi \quad , \tag{19}$$

where the last term may be called the interference term.

For the special case where $a_1 = a_2$,

$$I'_{2c} = 2I_1(1 + \cos\phi) \quad , \tag{20}$$

and I'_c ranges from 4 I_1 to 0, depending upon ϕ. It may be noted from eq. (19) that the average value of intensity over all phases is I_1+I_2.

In incoherent sources the interference term averages out and there is no interference.

Thus,

$$I_{2i} = I_1 + I_2 \quad , \tag{21}$$

$$= a_1^2 + a_2^2 \ . \tag{22}$$

The above results may be generalized for any number of waves by summation:

$$I_c = (\sum_k \vec{u}_k)(\sum_k \vec{u}_k{}^*) = \left|\sum \vec{u}_k\right|^2 \quad , \tag{23}$$

$$I_i = \sum_k (\vec{u}_k \vec{u}_k{}^*) = \sum_k |\vec{u}_k|^2 \quad . \tag{24}$$

Equations (23) and (24) state that interference occurs for coherent, but not for incoherent beams.

To summarize, coherent light is linear in amplitude only, while incoherent light is linear in intensity only. In other words, to find the resultant intensity of coherent light beams add the individual amplitudes first and then square the sum; to find the resultant intensity of incoherent light beams, square the individual amplitudes, then add them together.

Interference between light beams requires that they be correlated, and an improved understanding of the phenomenon can be obtained from a consideration of simple coherence theory, which is a branch of mathematics that deals with the statistics of correlations. Although coherence is a measure of the correlation between beams of light, the term has been extended to apply also to a source, which is said to be coherent if all beams from it are highly correlated.

Two types of coherence are delineated, temporal and spatial, and a Michelson interferometer may exhibit both. Temporal coherence is a measure of the ability of two similar beams which emanate from a common point to interfere; it is dependent on the retardation, δ, or alternatively, the (time) delay $\tau \equiv \delta/c$, where c is the speed of light. Temporal coherence is governed by an uncertainty relationship,

$$\Delta\tau\Delta\nu \sim 1$$

where the time $\Delta\tau$ is called the coherence time of the radiation and $\Delta\nu = \nu_1 - \nu_2$ is the (optical frequency) bandwidth of the radiation; this shows the temporal coherence to increase as the bandwidth decreases.

Spatial coherence is a measure of the ability of two similar beams emanating from separate points and having zero delay, to interfere; it relates to the finite size, or extension, of an actual source (a point being an idealization).

In a Michelson interferometer an extended source gives rise to two separated, but overlapping images in the plane of observation of the interference - one from each beam. Increasing the extension increases the spatial separation, the consequence of which is to decrease the spatial coherence. Thus limitations are placed on the size of the source or its defining aperture. If the source is an effective point, both images coincide spatially and only temporal coherence is relevant.

In the general case of two beams with both separation and delay, coherence depends on both the spatial separation and the delay, and the degree of coherence may range from fully coherent to incoherent, with partial coherence in between. The measured intensity of the light observed at the interference plane will of course depend on the joint coherence.

No significant interference can result from two beams which come from two separate, independent incoherent sources: the radiation from each incoherent source independently fluctuates so rapidly that constant phase difference cannot be maintained for a measurable duration. The interferometer, however, synchronizes two beams of nonmonochromatic radiation by deriving them from a common luminous origin, thus making interference feasible.

Polychromatic radiation may be regarded as made up of a continuum of monochromatic waves. From the uncertainty relationship of temporally coherent light, the quasimonochromatic components in the two beams will interfere; but the coherence times for the other, or mixed, components will be negligibly short. When the arms of the interferometer are balanced--i.e., at the null--<u>all</u> pairs of quasimonochromatic components will be effectively in phase ($\delta \sim 0$) and the intensity will be at the maximum. As the retardation is increased these component pairs will de-phase by varying degrees and the intensity will be reduced, until at some appropriate retardation, some particular component pair will again come into phase, causing the intensity to increase. Thus, as the retardation is varied, the intensity at the detector will consist of an alternating component owing to light of nearly the same wave number and a steady component owing to light of different wave numbers. The detector circuit responds just to the alternating component, generally referred to as the interferogram.

The reader may recall from the preceding chapter that if the source of the Michelson interferometer is coherent, the interferogram if expressed as a function of retardation will be a cosinusoid of constant modulus; cf Fig. 16(a). If, instead, the source were incoherent, the interferogram would be a modified cosinusoid with the modulus variable.

13. DERIVATION OF THE BASIC EQUATION OF FTS

The basic equation of Fourier transform spectroscopy shall now be derived for an ideal Michelson interferometer as described above, except the assumed source of radiation shall be polychromatic, as required for practical FTS.

Consider first the case of the two beams of monochromatic light $f_1(x)$ and $f_2(x)$+ which recombine at the beam splitter, as depicted in Fig. 20. The amplitudes, a, and wave numbers, ν_0, are equal, and the optical path difference between the two waves is δ. It is convenient to assume that one wave function is not shifted; then the shift of the other wave function $x_0 = \delta$. And because this is an LSI system, the resultant wave function is simply the sum of the individual wave functions. Thus,

---

+Designation of coordinates is arbitrary. Here we choose to express the wave function in its more familiar form.

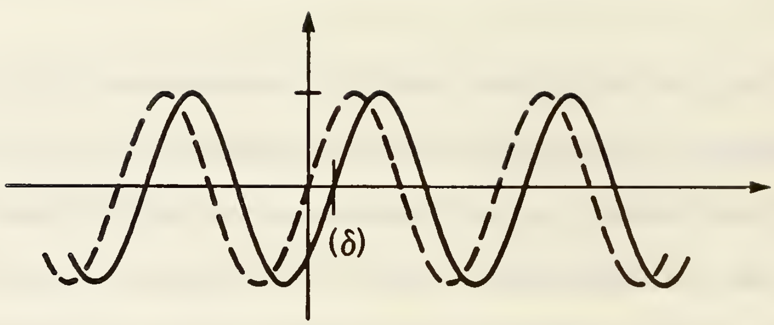


Figure 20. Two sinusoids of equal frequency and amplitude traveling in the same direction with constant phase difference. δ is the optical path difference.

$$f_{R,m}(x) = f_1(x) + f_2(x) = ae^{j2\pi\nu_0 x} + ae^{j2\pi\nu_0 (x-\delta)} \quad , \tag{25}$$

$$= a(1+e^{-j2\pi\nu_0\delta})e^{j2\pi\nu_0 x} \quad . \tag{26}$$

Now assume that the source of radiation is polychromatic. Again, because we are dealing with a LSI system, the resultant wave function for white light of all frequencies is simply the sum of all the individual wave functions corresponding to all of the different wave numbers which comprise the white light. Thus,

$$f_R(x) = \int_{-\infty}^{\infty} f(x)d\nu = \int_{-\infty}^{\infty} [A(\nu)(1+e^{-j2\pi\nu\delta})]e^{j2\pi\nu x}d\nu. \tag{27}$$

To simplify the mathematics we define a complex amplitude equal to the terms within the brackets; i.e.,

$$A_R(\nu,\delta) \equiv A(\nu)(1+e^{-j2\pi\nu\delta}) \ . \tag{28}$$

Then,

$$f_R(x) = \int_{-\infty}^{\infty} A_R(\nu,\delta)e^{j2\pi\nu x} \, d\nu. \tag{29}$$

We need to find an expression for the intensity, so recalling that intensity is given by the product of the complex amplitude and its complex conjugate,

$$I_R'(\nu,\delta) = A_R(\nu,\delta)A_R^*(\nu,\delta) \quad . \tag{30}$$

Carrying out the multiplication and using the definition of the cosine+ gives

$$I_R'(\nu,\delta) = 2A^2(\nu)[1+\cos(2\pi\nu\delta)]. \tag{31}$$

A new expression for intensity as a function of only a single variable, specifically, retardation, can be written by integrating all the intensities of different wave numbers.

$$I_R(\delta) \equiv \frac{1}{\bar{\nu}} \int_0^\infty I_R'(\nu,\delta)d\nu \quad , \tag{32}$$

where $\bar{\nu}$ is the mean wave number (spatial frequency).

+$\cos\alpha = \dfrac{e^{j\alpha} + e^{-j\alpha}}{2}$

Substituting $I_R'(\nu,\delta)$ from eq. (31) into eq. (32) gives

$$I_R(\delta) = 2\left(\frac{1}{\bar{\nu}}\right)\left\{\int_0^\infty A^2(\nu)d\nu + \int_0^\infty A^2(\nu)\cos(2\pi\nu\delta)d\nu\right\} . \qquad (33)$$

It may be shown⁺ that

$$2\left(\frac{1}{\bar{\nu}}\right)\int_0^\infty A^2(\nu)d\nu = \frac{1}{2}\, I(0) \quad . \qquad (34)$$

where $I_R(\delta=0) \equiv I(0)$.

Then,

$$I(\delta) = \frac{1}{2}\, I(0) + 2\left(\frac{1}{\bar{\nu}}\right)\int_0^\infty A^2(\nu)\cos(2\pi\nu\delta)d\nu . \qquad (35)$$

where the subscript has been dropped from I_R;

⁺In eq. (33), for $\delta\to\infty$, the cosine term oscillates very rapidly and therefore averages to zero. Thus, $I_R(\infty)=2\left(\frac{1}{\bar{\nu}}\right)\int_0^\infty A^2(\nu)d\nu = \frac{1}{2}\, I(0)$.

$$I(\delta) - \frac{1}{2} I(0) = 2\left(\frac{1}{\bar{\nu}}\right) \int_0^{\infty} A^2(\nu)\cos(2\pi\nu\delta)d\nu. \quad (36)$$

This equation may be recognized as a cosine-form of the Fourier transform equation+, cf eq. (9), with f(x) and $F^2(\xi)$, transform pairs. Thus, eq. (36) may be inverted [cf eq.(8)] to obtain

$$2\int_0^{\infty}[I(\delta) - \frac{1}{2} I(0)]\cos(2\pi\nu\delta)d\delta = \frac{1}{\bar{\nu}} A^2(\nu); \quad (37)$$

$$= I(\nu) \; ; \quad (38)$$

or,

$$I(\nu) = \int_{-\infty}^{\infty}[I(\delta) - \frac{1}{2} I(0)]\cos(2\pi\nu\delta)d\delta. \quad (39)$$

Equation (39) is the basic equation of FTS, and it states that the intensity at a single given wave number (ν) is the Fourier transform of the interferogram [$I(\delta) - \frac{1}{2} I(0)$, where δ is the retardation]. To obtain the spectra, the calculation is repeated, using eq. (39), for each wave number of concern. It may be noted that the interferogram consists of an alternating component, $I(\delta)$, and a steady component, $\frac{1}{2} I(0)$, as predicted in the chapter on Coherence and Interference.

+For an even function, $e(\alpha)$, $2\int_0^{\infty} e(\alpha)d\alpha = \int_{-\infty}^{\infty} e(\alpha)d\alpha$.

14. APODIZATION

Apodization (from the Greek) literally means the process of removing feet or the state of having no feet. In engineering jargon, these so-called feet would be referred to as side lobes.

Side lobes ineluctably arise in spectra owing to the finite extent, or spatial bandwidth, of actual interferograms which is imposed by the maximum excursion of the movable mirror; if L is the maximum optical path difference thus produced, the spatial bandwidth will be given by 2L. Usually the presence of these side lobes, which are artifacts of some prominence, is objectionable because it can produce spurious results. Thus, although the side lobes cannot be eliminated, they may be substantially reduced through apodization, which is obtained instrumentally or mathematically; e.g., the use of low-pass amplitude filters of various frequency response characteristics, of which the most common is the triangular shaped. This chapter will contrast the spectra of monochromatic light waves of infinite spatial bandwidth with those of finite spatial bandwidth, with and without apodization.

Spectra of monochromatic radiation of infinite spatial bandwidth

This case is an idealization because infinite bandwidths cannot be realized experimentally. Nevertheless, the case is of pedagogic value.

As shown in the preceeding section, the spectrum is simply the Fourier transform of the interferogram; e.g.

$$I(\xi) = \mathcal{F}\{I(x)\} \quad ,$$

where the interferogram is considered to be just the modulated component.

Monochromatic light waves may be represented by a cosine function of the generalized form, $\cos(n\pi x)$, where n is an integer. Thus, the spectrum is given by

$$\mathcal{F}\{\cos(n\pi x)\} = \frac{1}{2}\,\delta(\xi + \frac{n}{2}) + \frac{1}{2}\delta(\xi - \frac{n}{2}), \tag{40}$$

where δ denotes the delta function. For the special case of $\cos(\pi x)$, the spectrum is defined as the even impulse pair,

$$\delta\delta(\xi) \equiv \frac{1}{2}\,\delta(\xi + \frac{1}{2}) + \frac{1}{2}\,\delta(\xi - \frac{1}{2}) \quad . \tag{41}$$

This spectrum, as defined, is just a pair of delta functions of amplitude = 1/2, located at $\xi = \pm 1/2$.

For the generalized cosine function, $\cos(n\pi x)$, where n=2,3,..., the Fourier transform, or spectrum, is just a shifted form of even impulse pair which is located at $\xi = \pm\frac{1}{2}$, with the amplitude of each impulse remaining unchanged at $\frac{1}{2}$. Figure 21 depicts various cosine functions and their spectra, and the spectra are seen to diverge correspondingly with expansion of the cosinusoid.

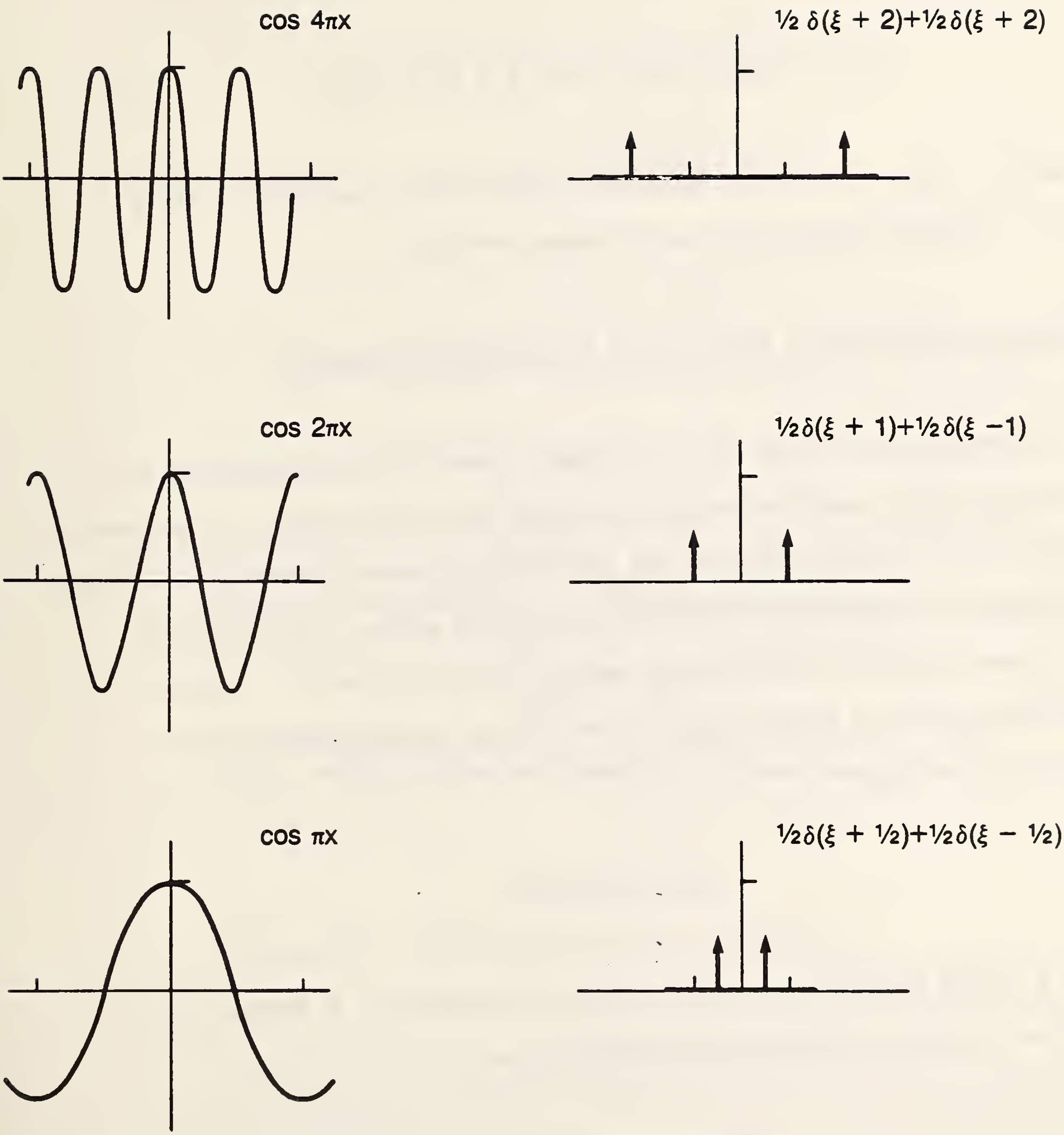


Figure 21. Expansion of a cosinusoid and corresponding shifts in its spectrum.

When convolved with a function, f(x), the even impluse pair has a duplicating property combined with a one-half reduction property; i.e.,

$$\delta\delta(x)*f(x) = \frac{1}{2}\, f(x+\tfrac{1}{2}) + \frac{1}{2}\, f(x-\tfrac{1}{2}) \quad . \qquad (42)$$

Figure 22 illustrates the convolution of $\delta\delta(x)$ with an arbitrary function, f(x). A shifted impulse pair would behave similarly.

Spectra of monochromatic radiation of finite spatial bandwith

A common type of filter is the "boxcar" (named for its resemblance to a railroad boxcar), described mathematically by the rect function. Components falling within the boundaries of the rectangle are passed unattenuated, while components falling outside the boundaries are not passed; such filtering occurs when the mirror excursion is the sole filtering factor. Thus, an unattenuated wave of finite spatial bandwidth may be represented by an equation of the form

$$\cos(n\pi x)\mathrm{rect}(\tfrac{x}{b}) \; ,$$

where b is the bandwidth, or width of the rect function. As before, the spectrum is given by the Fourier transform; i.e.,

$$\mathcal{F}\{\cos(n\pi x)\mathrm{rect}(\tfrac{x}{b})\} \quad .$$

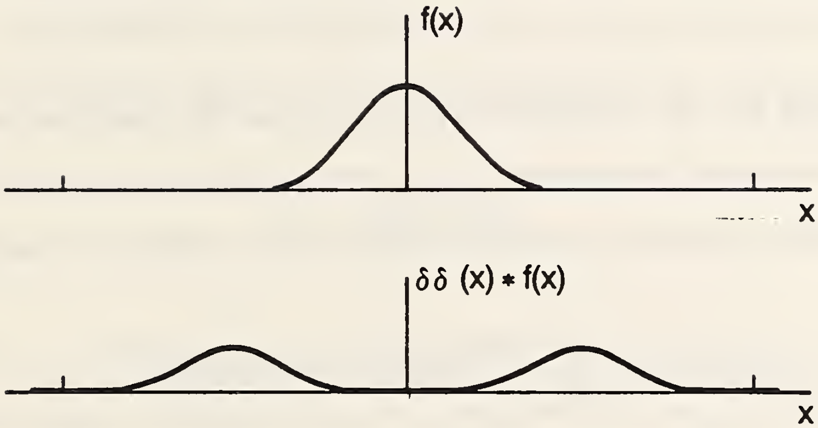


Figure 22. The convolution $\delta\delta(x)$ with $f(x)$.

Making use of the transform property of a product (see Properties of Fourier Transforms), eq. (40), and the similarity theorem+:

$$\mathcal{F}\{\cos(n\pi x)\mathrm{rect}(\frac{x}{b})\} = \mathcal{F}\{\cos(n\pi x)\} * \mathcal{F}\{\mathrm{rect}(\frac{x}{b})\}$$

$$=[\frac{1}{2}\delta(\xi+\frac{n}{2}) + \frac{1}{2}\delta(\xi-\frac{n}{2})]*b\ \mathrm{sinc}\ b\xi. \qquad (43)$$

Thus, the spectrum of a cosinusoid of finite spatial bandwidth is a pair of scaled sinc functions, as shown in the example of Fig. 23, where n=2 and b=4 were chosen. Note the prominent side lobes, particularly those of negative polarity.

These side lobes may be reduced substantially and the negative portions eliminated by apodization, specifically by substituting a tri function for the rect function, while maintaining constant bandwidth. Then, the spectrum is given by

$$\mathcal{F}\{\cos(n\pi x)\mathrm{tri}(\frac{x}{b})\} = \mathcal{F}\{\cos(n\pi x)\} * \mathcal{F}\{\mathrm{tri}(\frac{x}{b})\}$$

$$=[\frac{1}{2}\delta(\xi+\frac{n}{2} +\frac{1}{2}\delta(\xi\frac{n}{2})]*b\ \mathrm{sinc}^2 b\xi. \qquad (44)$$

+Similarity theorem: If $\mathcal{F}\{(x)\}=F(\xi)$, then $\mathcal{F}\{f(ax)\} = \frac{1}{|a|}F(\frac{\xi}{a})$.

To correspond to the previous example, let n=2; b=2. Then, the spectrum is given by

$$\mathcal{F}\{\cos(2\pi x)\mathrm{tri}(\tfrac{x}{2})\}=[\tfrac{i}{2}\ \delta(\xi+1)+\tfrac{i}{2}\ \delta(\xi-1]*2\ \mathrm{sinc}^2 2\xi. \qquad (45)$$

$$= \mathrm{sinc}^2 2(\xi+1)+\mathrm{sinc}^2 2(\xi-1). \qquad (46)$$

Figure 23 depicts this triangularly apodized spectrum together with its interferogram, as well as the unapodized spectrum and its interferogram, previously discussed, for comparison.

Although apodization reduces the side lobes, there is a trade-off with the resolution. The degradation of resolution with apodization, as well as with decreasing interferogram bandwidth, will be demonstrated in the next chapter.

15. RESOLUTION

Resolution, as defined here, is the capability of an optical system (including the means of observation) of rendering distinguishable a pair of closely-adjacent, parallel lines of equal intensity.

When the lines are far apart, each is readily distinguishable, however, as they approach one another, a critical separation is reached beyond which the lines appear to merge. The limit of resolution is related to the critical separation, and it is customary to measure the separation from the center of one vertical line to the center of the other vertical line. Thus, on purely geometrical grounds, the limit of resolution will be lower for narrow lines than for wide lines.

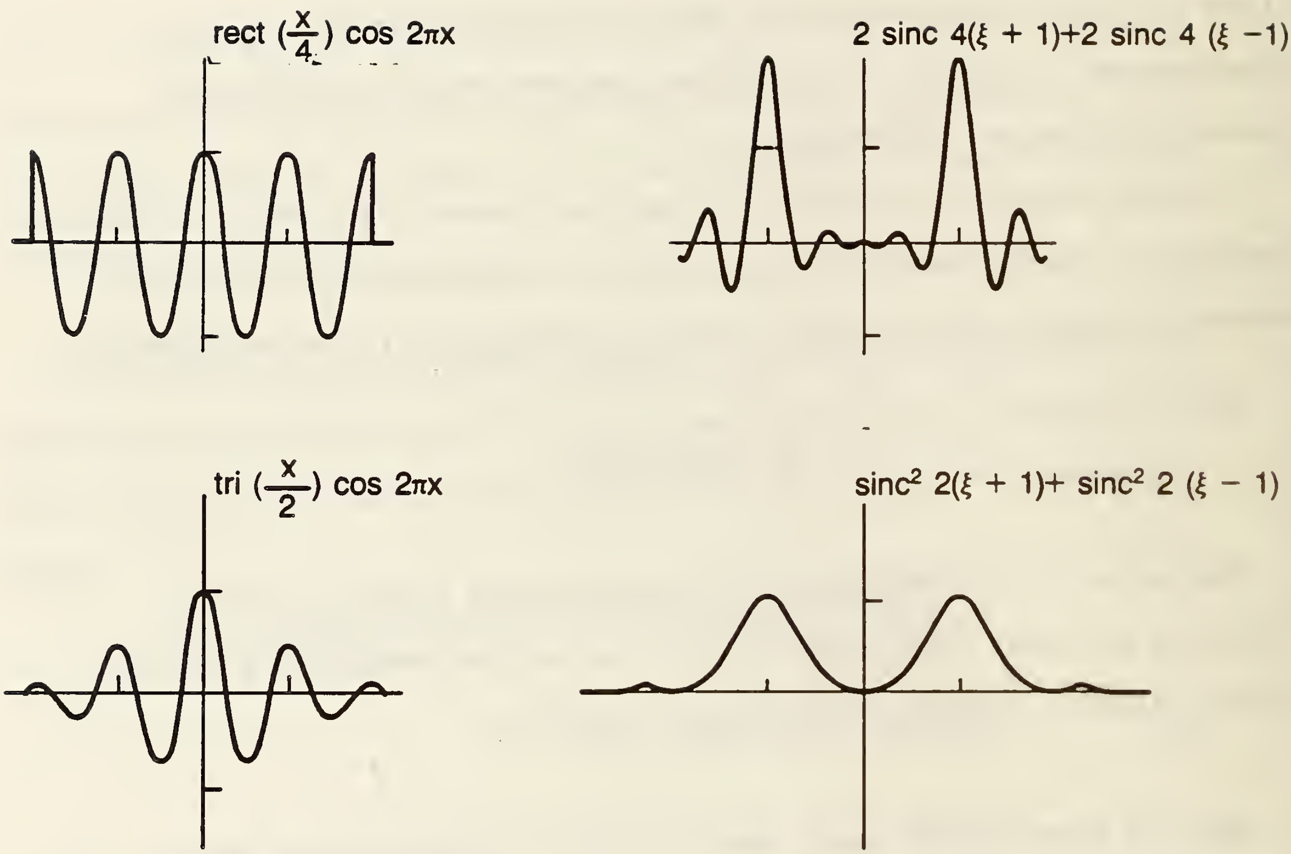


Figure 23. Spectra of monochromatic radiation of finite bandwidth. Upper right: unapodized spectrum; lower right; apodized spectrum. The corresponding interferograms are to the left.

Actually, resolution is a complex phenomenon which depends on many variables such as wavelength, coherence/incoherence, and (for coherent light only) phase difference. Because resolution is of practical importance, some measures of the limit of resolution are often given for convenience. These measures are arbitrarily derived, however, and generally should be considered merely as rough approximations which may also serve for purposes of comparison or prediction.

The resolution limit is commonly expressed in terms of either 1) half-width (by which is meant the full width of the line at one-half the peak intensity) or 2) the Rayleigh criterion. These criteria yield roughly comparable results and are illustrated in Figs. 24 and 25.

Figure 24 depicts three different views of the same line image. Figure 24(a) shows the geometrical rectangular shape; (b) shows schematically the distribution of intensity in two dimensions: the intensity is a maximum at the center (x=0) and decreases rapidly toward the edges; and (c) shows the distribution of intensity in the common one-dimensional format.

The Rayleigh criterion for the limit of resolution was postulated originally in connection with prism and grating spectroscopes where the intensity distribution for monochromatic light is of the form of a sinc^2 function. The minimum resolvable separation was assumed to occur when the peak of one sinc^2 function coincides with the first zero of the other sinc^2 function, as shown in Fig. 25. The Rayleigh separation is indicated by RC, and for comparison, the half-width is indicated by HW. The limit of resolution may be expresed in terms of phase difference or wave number. For

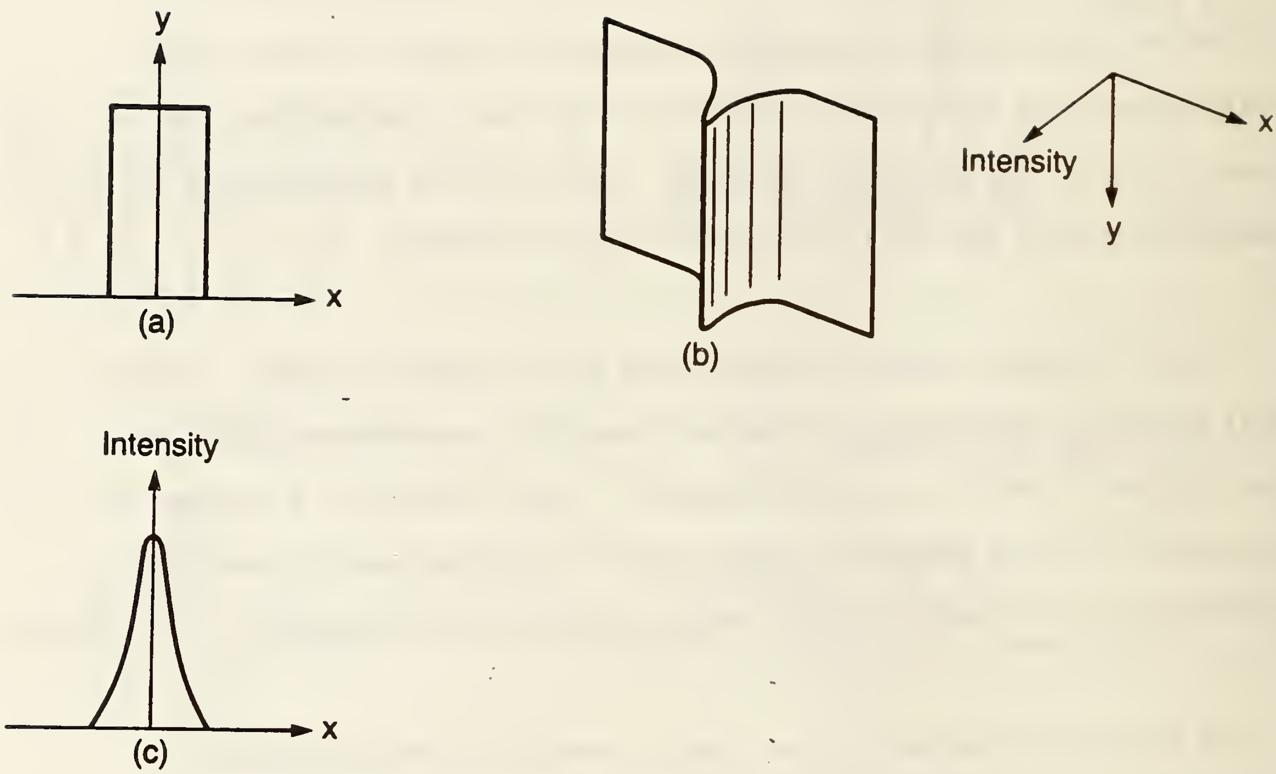


Figure 24. Three different views of the same line image: (a) shows the geometrical rectangular shape; (b) show schematically the distribution of intensity in two dimensions; (c) shows the distribution of intensity in one dimension.

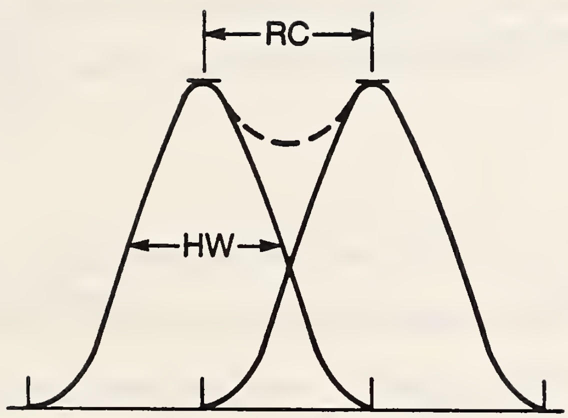


Figure 25. Limit of resolution according to the Rayleigh criterion for a pair of sinc^2 functions; RC indicates the separation. HW, the half-width of one function, is shown for comparison.

want of a better criterion Rayleigh's has been commonly adopted for resolution problems in general.

Now two important properties of Fourier transform spectroscopy relating to resolution can be proven: (1) that apodization degrades the resolution; (2) that the limit of resolution decreases with increasing interferogram bandwidth, or in other words the maximum optical path difference.

Actually, (1) has already been tacitly demonstrated; please refer to Fig. 23. The half-width of the sinc function (unapodized case) is observed to be less than the half-width of the $sinc^2$ function (apodized case). Thus, by the half-width criterion of resolution, the unapodized spectrum is of higher, or better, resolution (i.e., the limit of resolution is lower) than the apodized spectrum.

For (2), refer to Fig. 26, which shows two monochromatic waves of the same frequency, but having different interferogram bandwidths (one twice the other), and their corresponding spectra. Note that the half-width of the spectrum corresponding to the wider interferogram bandwidth is less than the half-width of the other spectrum. Thus, the limit of resolution of the spectrum is less, the wider the interogram bandwidth. Indeed, it may be shown, either geometrically or analytically, that the resolution varies inversely as the interferogram bandwidth, or maximum optical path difference (see Bell, Chapt. 6, cited in the Bibliography).

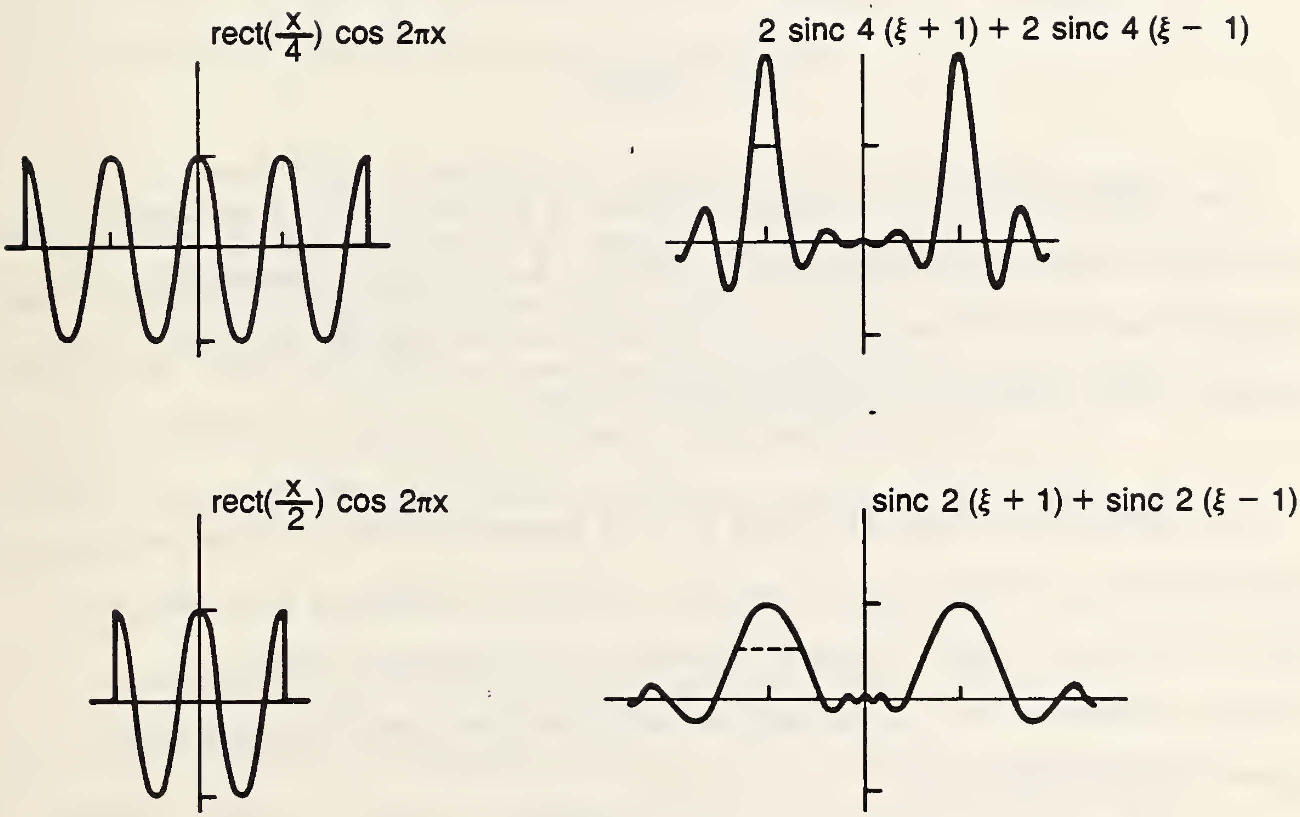


Figure 26. Showing that contraction of the interferogram bandwidth (left) causes the limit of resolution of the spectrum to increase (right). (The half-width is a measure of resolution limit).

The spectra which have been shown in this chapter relate to the curtailment or filtering of monochromatic light and are purely instrumental in nature; hence they are referred to as instrumental line shapes (ILS). That is, these spectra will occur with and without a test sample in place, owing to the interaction of light waves and interferometer. Hence these spectra are artifacts which must be removed when test samples are examined in the instrument. This is accomplished by the mathematical process of deconvolution, to be illustrated in a later chapter.

16. SAMPLING

The interferogram is a continuous record of signal, which in mathematical terms constitutes a one-dimensional function. But, for data processing by a digital computer it is necessary to represent the function by an array of its sampled values taken on a discrete set of points.

One might think intuitively that for the sampled data to be an accurate representation of the original function, it would be necessary to sample the data at extremely close intervals, however, for a particular class of function, common in FTS, it is required only that the sampling interval not exceed a determinable limit.

The particular class of function of concern is termed a band-limited function, defined as one whose Fourier transform has nonzero values only within a finite region, or band, of the frequency space. Figure 27 is an illustration of a band-limited function. Here, f(x) is the band-limited function, F(ξ) is its Fourier transform, and W is the spectral bandwidth.

If sampling is made in accordance with the above criteria, it is quite feasible to reconstruct approximately the original function from the sampled function. Otherwise, irreconcilable artifacts generally would result. Besides being a tremendous time saver and a convenience, in many cases sampling makes the required data processing feasible.

Various methods may be employed for sampling and the intervals between sample points are not required to be regular. Notwithstanding, comb-function sampling is the simplest method and for FTS, also the most appropriate. The sampled (and digitized) data are commonly processed with the use of a fast Fourier transform algorithm (FFT) to reduce data processing time still further. Use of the FFT algorithm requires that the sampling be precisely periodic.

For comb-fuction sampling the prescribed sampling interval must be no larger than the spectral bandwidth, W(cf Fig. 27). Hence, the sampling rate or frequency, which is just the reciprocal of the sampling interval, must be no smaller than the spectral bandwidth, W. The critical sampling interval and the critical sampling rate/frequency, are commonly referred to as the Nyquist interval and the Nyquist rate/frequency.

We proceed now to consider three cases of comb-function sampling and their associated spectra. The first two cases treat idealized comb-function

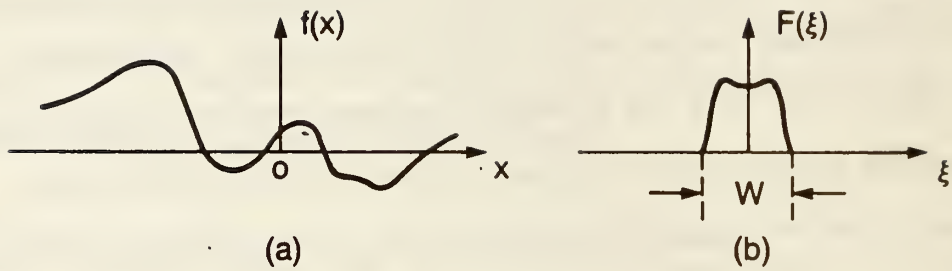


Figure 27. A band-limited function (a) and its Fourier transform (b); W is the spectral bandwidth (after Gaskill).

sampling, in which the comb function is of infinite extent; the last, with real comb-function sampling, in which the comb function is abridged.

Idealized comb-function sampling, proper sampling

The comb function has been described previously as an array of unit-area δ-functions, centered at the origin and spaced one unit apart. That is,

$$\mathrm{comb}(x)=\sum_{n=-\infty}^{\infty}\delta(x-n)\ , \tag{47}$$

where n is an integer. Figure 28 is a graphic representation of the comb function.

The scaled version of the comb function is

$$\frac{1}{|b|}\ \mathrm{comb}\left(\frac{x}{b}\right)=\sum_{n=-\infty}^{\infty}\delta(x-nb), \tag{48}$$

which is just an array of unit-area δ-functions, centered at the origin and spaced |b| units apart. Figure 29 depicts the scaled comb function.

Consider the band-limited function f(x) shown in Fig. 27. The sampled function $f_s(x)$ is given by the product of f(x) and the scaled comb-sampling function; i.e.,

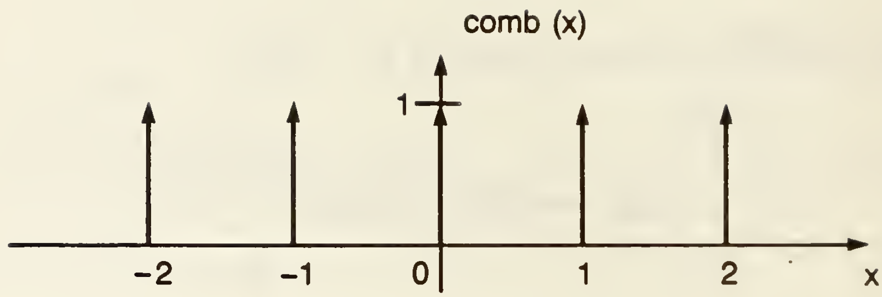


Figure 28. Graphical representation of the comb function.

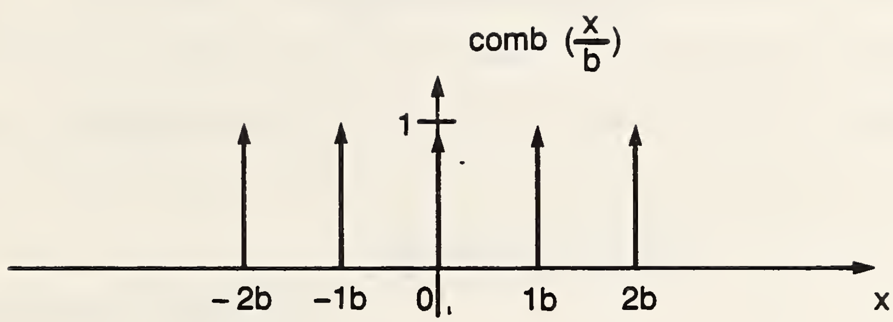


Figure 29. Scaled comb function.

$$f_s(x)=f(x)\left[\frac{1}{x_s}\,\mathrm{comb}\left(\frac{x}{x_s}\right)\right] \tag{49}$$

$$=\sum_{n=-\infty}^{\infty} f(nx_s)\,\delta(x-nx_s)\ , \tag{50}$$

where x_s = the Nyquist sampling interval.

The spectrum of $f_s(x)$ is

$$F_s(\xi)=\mathcal{F}\{f_s(x)\}=\mathcal{F}\left\{f(x)\ \left[\frac{1}{x_s}\,\mathrm{comb}\left(\frac{x}{x_s}\right)\right]\right\} \tag{51}$$

$$\mathcal{F}\{f(\tfrac{x}{b})\}=|b|F(b\xi)$$

$$F_s(\xi)=F(\xi)*\mathrm{comb}(x_s\xi) \tag{52}$$

Since $\xi_s\equiv x_s^{-1}$,

$$F_s(\xi)\ =F(\xi)*\mathrm{comb}\left(\frac{\xi}{\xi_s}\right) \tag{53}$$

and from (50)

$$F_s(\xi)=\xi_s\sum_{n=-\infty}^{\infty} F(\xi-n\xi_s)\ . \tag{54}$$

Thus, $F_s(\xi)$ is seen to consist of an array of functions, each having the form of $F(\xi)$(the Fourier transform of the original function) repeated at intervals of ξ_s along the spatial-frequency axis, as shown in Fig. 30. The various facsimiles of $F(\xi)$ are often referred to as the spectral orders of $f_s(x)$, with $F(\xi-n\xi_s)$ known as the nth spectral order.

Idealized comb-function sampling, undersampling

For an example consider, as before, the band-limited function f(x) of Fig. 27. If the sampling interval is larger than the critical value W, the spectral orders overlap one another, as shown in Fig. 31. Because of the overlap, $F_s(\xi)$ is no longer a series of distinct facsimilies of $F(\xi)$, but takes on additional values which are spurious. The condition of overlap is referred to as aliasing and the artifacts which result generally cannot be reconciled.

Real comb-function sampling

A real sampling function has to be abridged, rather than infinite as for the idealized cases, above. By means of a sequence of mathematical operations and their pictorial representations, Fig. 32 traces an example of real comb-function sampling, from interferogram to valid spectrum.

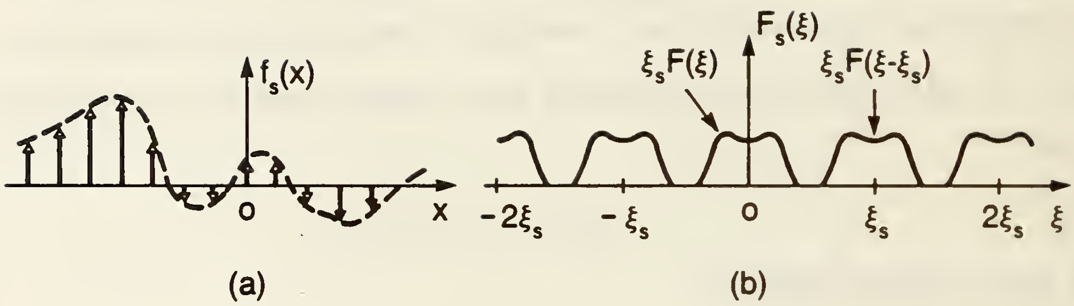


Figure 30. Comb-function sampling. (a) Sampled function, (b) Spectrum of sampled function (after Gaskill).

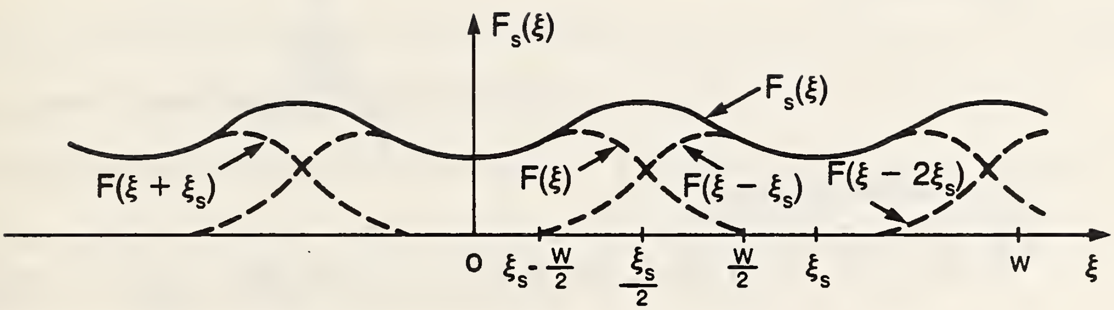
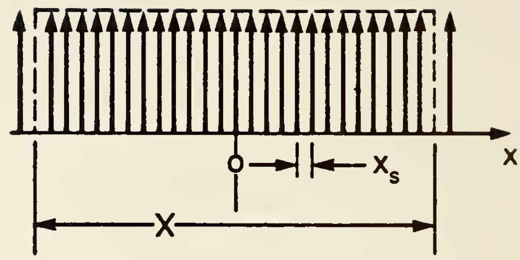


Figure 31. Spectrum of a sampled function when the sampling rate is less than the Nyquist rate.

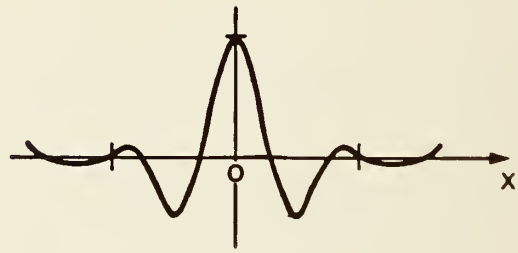
1. Sampling function

$$S(x) = \frac{1}{x_s}\,\mathrm{comb}\left(\frac{x}{x_s}\right)\cdot \mathrm{rect}\left(\frac{x}{X}\right)$$



2. Interferogram

$$f(x) = e^{-\pi x^2}\cdot \cos 2\pi x$$

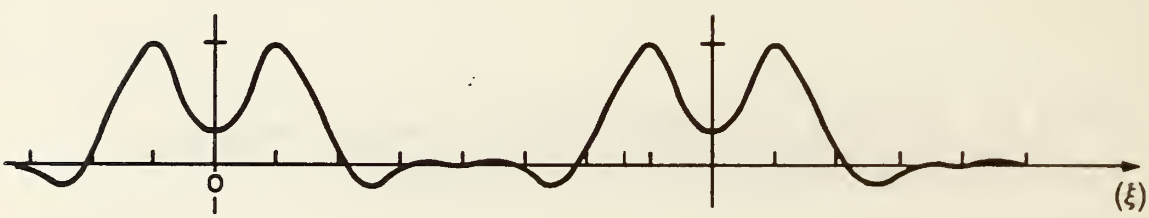


3. The sampled interferogram

$$f_s(x) = S(x)\cdot f(x)$$

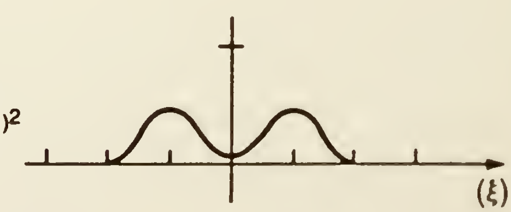
4. The sample spectrum

$$\mathcal{F}\left\{f_s(x)\right\} = \mathcal{F}\left\{S(x)\right\} * \mathcal{F}\left\{f(x)\right\} = [\mathrm{comb}\,(x_s\xi) * X\,\mathrm{sinc}\,(X\xi)] * [e^{-\pi\xi^2} * \delta\delta\,(\xi)]$$



5. The true spectrum

$$\mathcal{F}\left\{f(x)\right\} = \mathcal{F}\left\{e^{-\pi x^2}\right\} * \mathcal{F}\left\{\cos 2\pi x\right\}$$
$$= e^{-\pi\xi^2} * \delta\delta\,(\xi)$$
$$= \tfrac{1}{2}\,e^{-\pi(\xi+1)} + \tfrac{1}{2}\,e^{-\pi(\xi-1)^2}$$



6. The sampled spectrum, deconvoluted

$$\mathcal{F}\left\{f_s(x)\right\} * \mathcal{F}^{-1}\left\{S(x)\right\} = \mathcal{F}\left\{f(x)\right\} * \mathcal{F}\left\{S(x)\right\} * \mathcal{F}^{-1}\left\{S(x)\right\} = \mathcal{F}\left\{f(x)\right\}$$

Figure 32. Example of real comb-function sampling.

Figure 32(1) shows the sampling function S(x), taken to be an abridged comb function, with x_s, the sampling interval and X, the bandwidth; (2) shows the interferogram, f(x), assumed to be a gaussian-damped cosinusoid; (3) is the sampled interferogram, $f_s(x)$; (4) is the sampled spectrum, $\mathcal{F}\{f_s(x)\}$ (the Fourier transform of a gaussian function is also a gaussian function); (5) shows the true spectrum, $\mathcal{F}\{f(x)\}$-- note the discrepancy in amplitude as well as shape between the sampled spectrum and the true spectrum; (6) shows analytically that deconvolution of the sampled function with the sampling function yields the true spectrum.

17. ANALOG-TO-DIGITAL CONVERSION

The sampling and accompanying digitization are carried out by an analog-to-digital converter (A/D; ADC). Hence, the sampled analog data of the interferogram, which directly represents a measured quantity, is encoded into a "language", or digital code, which can be handled by a digital computer; a sequence of binary numbers, each element of which has the value "0" or "1". An example of a digital code is 101110101, composed of nine bits, or units of information.

All values of analog input are subdivided equally, or quantized, by partitioning the continuum into 2^n discrete ranges, where n is the number of bits. All analog values which fall within a given range ($1/2^n$) are represented by the same digital code, which corresponds to the nominal mid-range value. Thus, in the A/D conversion process, there exists an inherent quantization uncertainty which decreases as the number of bits increases. For example, for a converter with 20 bits and 10V full scale, the uncertainty would be (2^{-20}) $10V \approx 10\mu V$ or $\pm 5\mu V$.

In addition to quantization uncertainty, any errors in offset, gain, or linearity of the converter could lead to coding errors. For details of A/D conversion, see Sheingold, cited in the Bibliography.

18. SAMPLE INTERFEROGRAMS AND SPECTRA

This chapter presents a sampling of monochromatic interferograms and the corresponding spectra, some of which have been given previously; see Fig. 33. In such simple cases it is often possible by inspection of an interferogram to deduce its approximate spectrum, and vice versa.

That the radiation is monochromatic means that one function of the interferogram is a cosinusoid, whose spectrum is a scaled even impulse pair which will result in spectral twins. The other function(s) of the interferogram is some kind of truncation function; e.g., rect, tri, Gauss, or product thereof. Viewing an interferogram, which is a product, one conjectures which function,

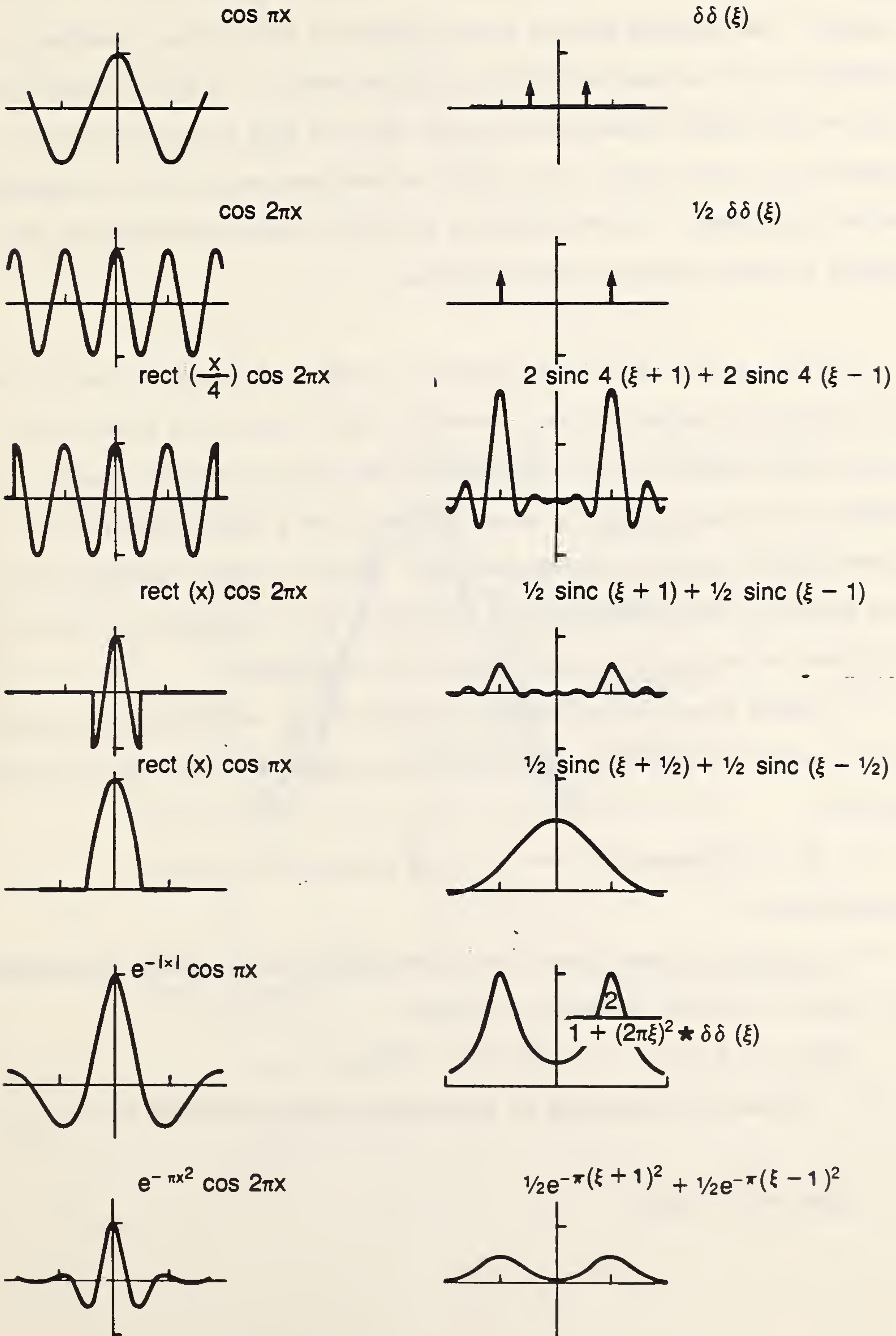


Figure 33. Sample monochromatic interferograms and their spectra. Small ticks show where the variable have a value of unity.

f_T, multiplied the cosinusoid. The spectrum would just be the transform of f_T (scaled). The converse process entails inferring the inverse transform of a spectrum, which includes multiplying by a cosinusoid. It may be noted from Fig. 34 that several common functions do not vary much in shape from one another; e.g., tri, Gauss, sinc, $sinc^2$, so one need not be overly concerned in making conjectures. Figure 35 shows a more complicated interferogram, the product of three functions rather than two.

Finally, in Fig. 36 we show examples of simple polychromatic spectra, and one can begin to appreciate the remakable capabilities of the computer to untangle what appears to be inextricable. And Fig. 37 shows an example of a complex polychromatic interferogram (obtained with a wedged beamsplitter interferometer) and its derived spectrum. Perhaps it seems incredible that what appears to be garbage or noise can yield such a distinct spectrum, yet it truly can and the analytic proof thereof is quite simple.

1) Assume polychromatic radiation consisting of infinite cosine waves of various optical frequencies, amplitudes, and phases, all traveling in the same direction.

2) By the Superposition Principle the waves add to give an interferogram.

3) The Fourier transform of the interferogram is its spectrum; thus apply the Linearity Theorem, or Addition Theorem:

$$\mathcal{F}\{f_1(x) + f_2(x) + \ldots\} = \mathcal{F}\{f_1(x)\} + \mathcal{F}\{f_2(x)\} + \ldots$$

4) Consider the transforms of the various cosine functions in 1); e.g.,

$$\mathcal{F}\{\cos(\pi x)\} = \delta\delta(x)$$

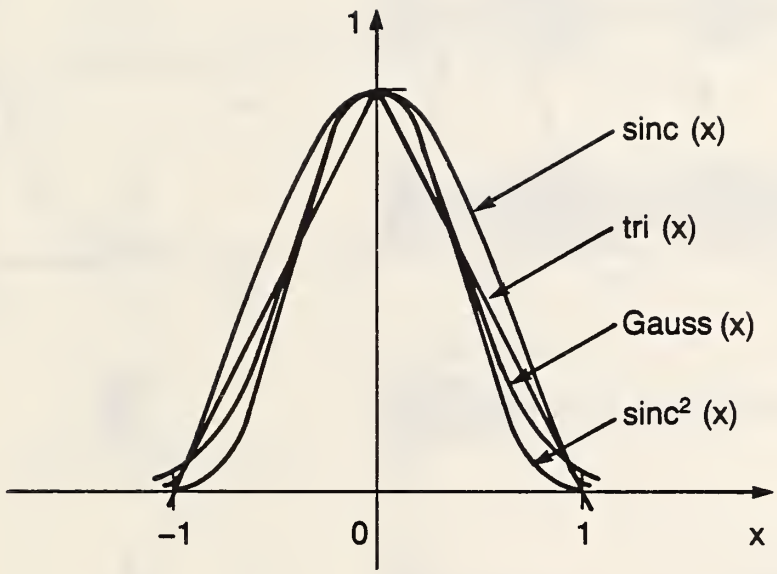


Figure 34. Similarities of the tri, sinc, sinc^2, and Gauss functions (after Gaskill).

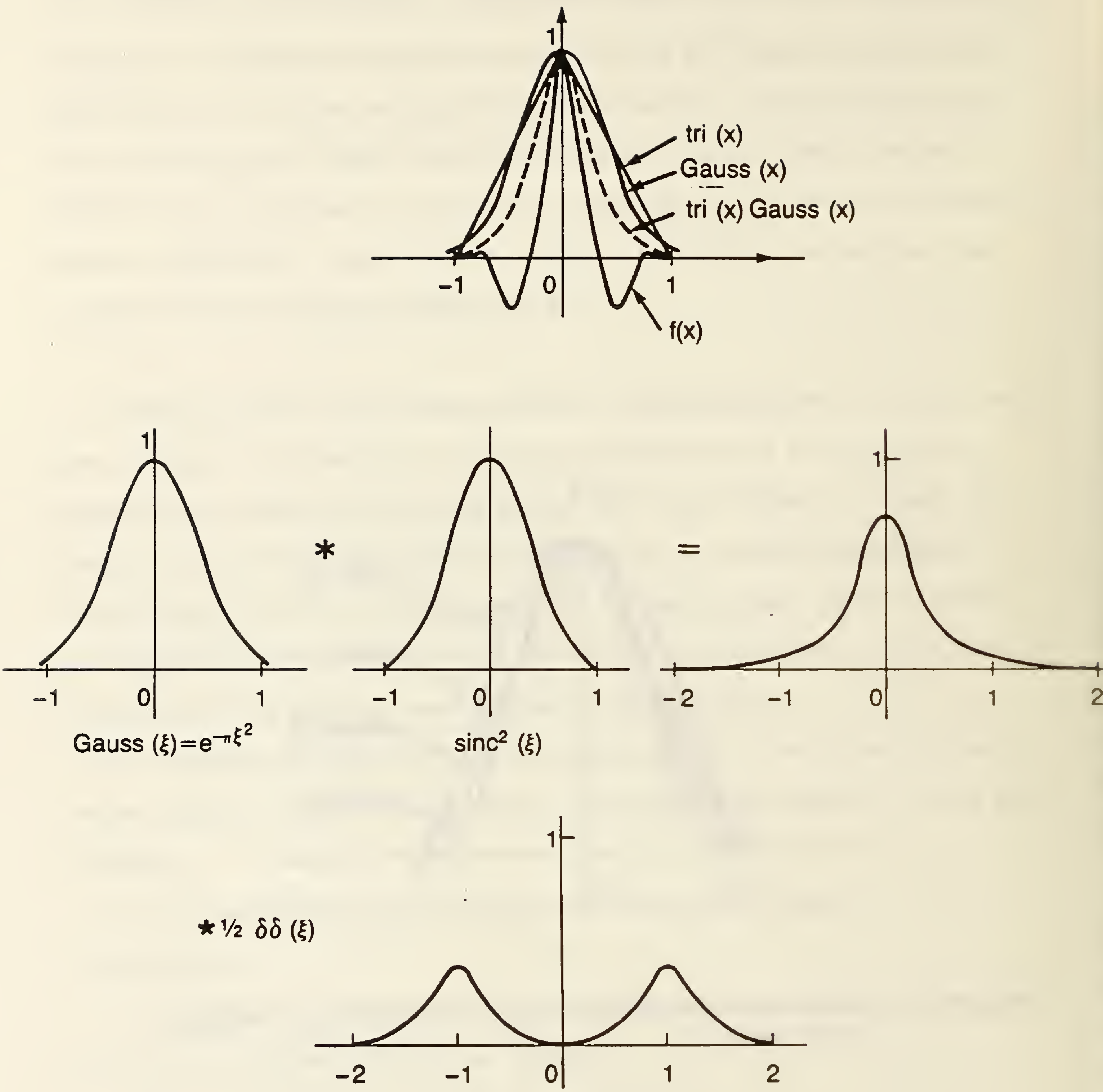


Figure 35. Spectrum of $f(x) = \text{Gauss}(x)\text{tri}(x)\cos\pi x$ by graphic construction; $\mathcal{F}\{f(x)\} = e^{-\pi\xi^2} * \text{sinc}^2(\xi) * 1/2\ \delta\delta(\xi)$.

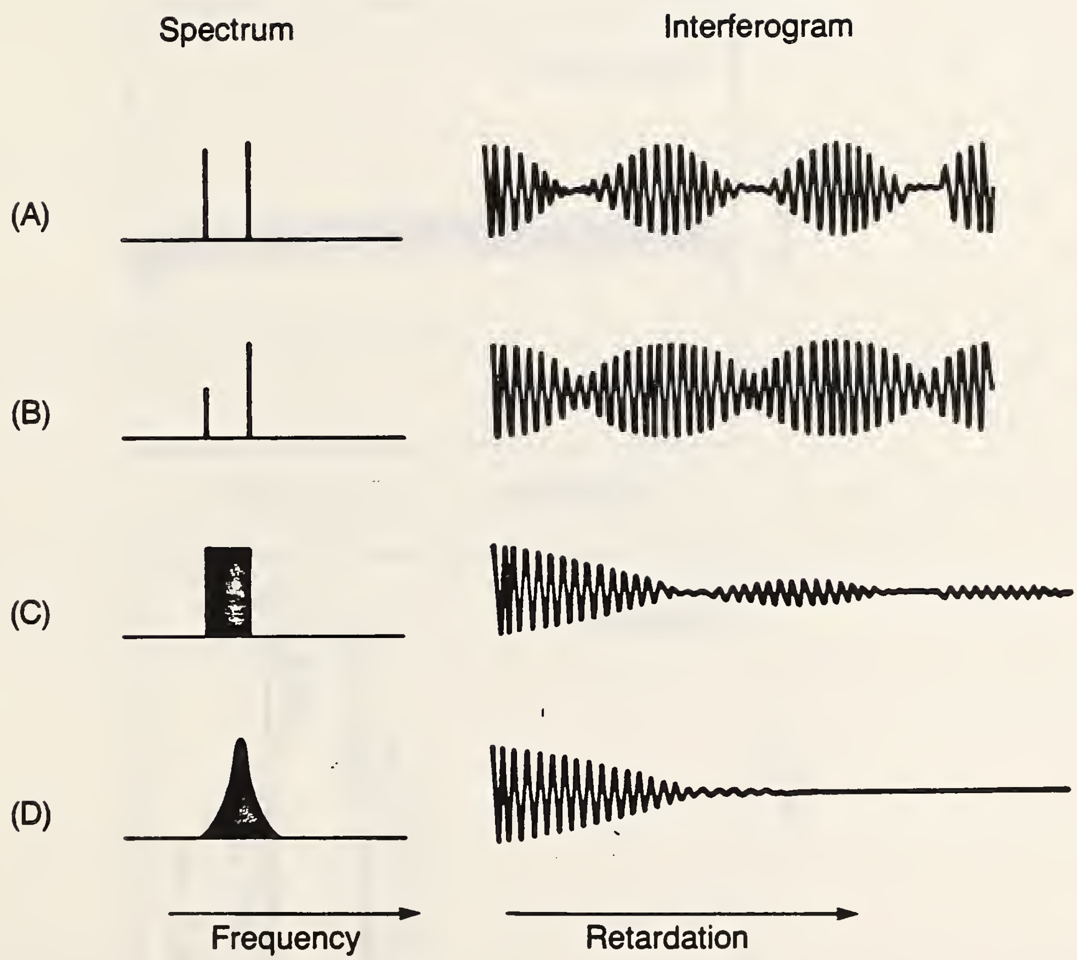


Figure 36. Simple polychromatic spectra and their interferograms (after Griffith).

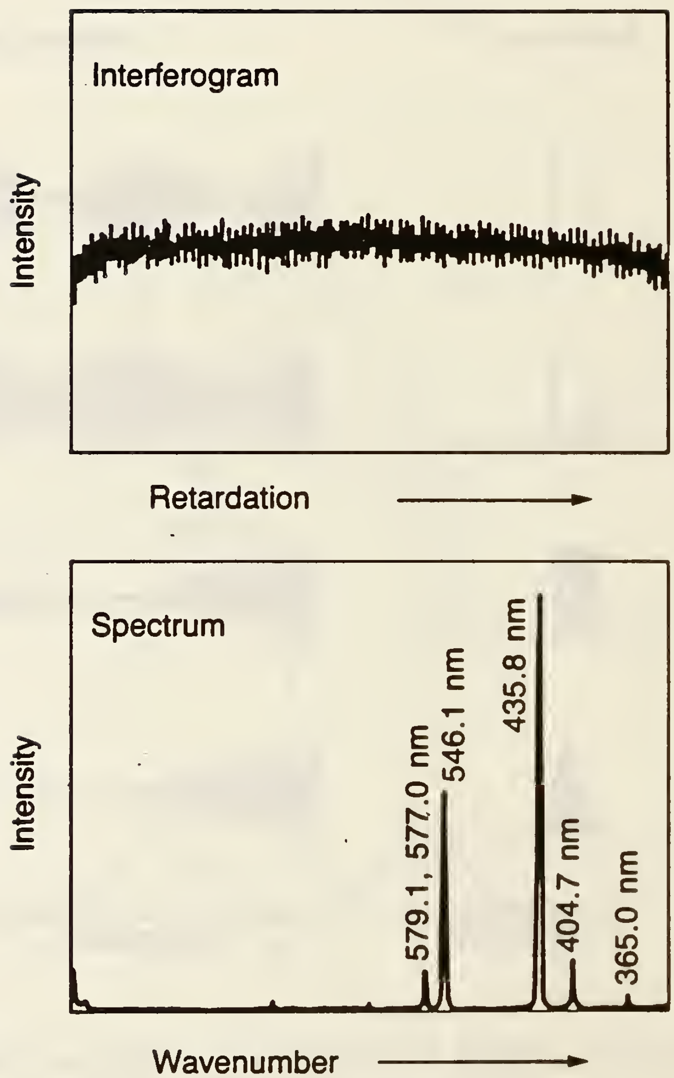


Figure 37. Interferogram and spectrum of a low pressure mercury lamp (after Okamoto, et al).

The amplitude spectrum of each cosine function is similar, being some form of an even impulse pair. The optical frequency of the cosinusoid determines the location of the impulse pair, whose spacing will vary inversely with the optical frequency of the cosinusoid. The amplitude of each impulse pair will be $A_i/2$, where A_i is the amplitude of the cosinusoid. Thus each wave unmber is uniquely determined, or in other words positioned on the spectrum.

BIBLIOGRAPHY

Bell, R.J., Introduction Fourier Transform Spectroscopy, Academic Press, New York, 1972.

Bracewell, R.N., The Fourier Transform and its Applications, second edition, McGraw-Hill, New York, 1978.

Burnett, C.R., Hirschberg, J.G., and Mack, J.E., "Diffraction and Interference", in Handbook of Physics, Condon, E.U. and Odishaw, H., eds., McGraw-Hill, New York, 1958.

Gaskill, J.D., Linear Systems, Fourier Transforms, and Optics, John Wiley and Sons, New York, 1978.

Goodman, J.W., Introduction to Fourier Optics, McGraw-Hill, New York, 1968.

Griffiths, P.R., Chemical Infrared Fourier Transform Spectroscopy, John Wiley and Sons, New York, 1975.

Okamoto, T., Kawata, S., and Minami, S., Applied Optics 23, 269 (1984).

Sheingold, D.H. (ed.), Analog-Digital Conversion Notes, Analog Devices, Inc., Norwood, MA 1980.

Steel, W.H., Interferometry, second edition, Cambridge University Press, Cambridge, 1983.

NBS-114A (REV. 2-80)

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET (See instructions)	1. PUBLICATION OR REPORT NO. NBSIR-86/3339	2. Performing Organ. Report No.	3. Publication Date MARCH 1986

4. TITLE AND SUBTITLE

Introduction to Fourier Transform Spectroscopy

5. AUTHOR(S)
Julius Cohen

6. PERFORMING ORGANIZATION (If joint or other than NBS, see instructions)

NATIONAL BUREAU OF STANDARDS
DEPARTMENT OF COMMERCE
WASHINGTON, D.C. 20234

7. Contract/Grant No.

8. Type of Report & Period Covered

9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS (Street, City, State, ZIP)

SAME

10. SUPPLEMENTARY NOTES

☐ Document describes a computer program; SF-185, FIPS Software Summary, is attached.

11. ABSTRACT (A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here)

This document is a simplified, concise introduction to Fourier Transform spectroscopy. The emphasis is on concepts and comprehension, and abundant diagrams are provided as an aid. The work is organized into three parts: first, a selective, but adequate review of Fourier transform mathematics, next, a treatment of the physics of a simple Michelson interferometer, and last, salient topics in Fourier transform spectroscopy.

12. KEY WORDS (Six to twelve entries; alphabetical order; capitalize only proper names; and separate key words by semicolons)
apodization; coherence; convolution; deconvolution; Fourier transforms; infrared; interferometry; phase; resolution; sampling; spectrophotometry

13. AVAILABILITY

[XX] Unlimited
☐ For Official Distribution. Do Not Release to NTIS
☐ Order From Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402.
[XX] Order From National Technical Information Service (NTIS), Springfield, VA. 22161

14. NO. OF PRINTED PAGES
101

15. Price
$16.95

USCOMM-DC 6043-P80