

CORRECTION OF PHASE ERRORS IN FOURIER SPECTROSCOPY*

C. D. Porter and D. B. Tanner†

*Department of Physics
The Ohio State University
Columbus, OH 43210*

Received July 22, 1982

The correction of phase error in Fourier transform spectroscopy is discussed. Phase error is corrected by multiplying the inverse transform of the measured interferogram by $e^{-i\phi(\omega)}$, where $\phi(\omega)$ is the phase error in the spectrum. This error is determined from a low-resolution, two-sided interferogram. It is absolutely essential that the truncation-apodization function which multiplies the interferogram have an even part which is symmetric about the $x=0$ position. If this function is not properly located, large errors in the corrected spectrum can occur.

Key words: Fourier transform spectroscopy, phase correction, interferometry.

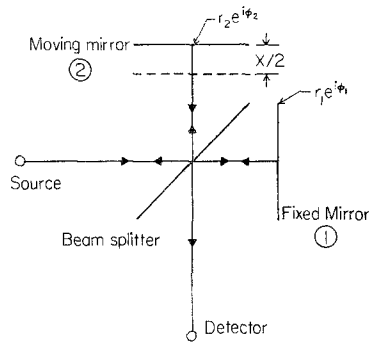
I. Introduction

This paper presents a discussion of a technique for correcting phase error, one of the errors which arise in Fourier transform spectroscopy. Phase error occurs for a variety of reasons: systematic error in the measurement of the optical path difference, misalignment of the interfer-

*Supported by the National Science Foundation through grant DMR-8110422.

†Current address: Physics Dept., University of Florida, Gainesville, FL 32611

ometer, or intrinsic asymmetry between the interferometer arms. These errors produce an interferogram which is asymmetric about its central maximum and whose Fourier transform is inherently a complex function of frequency. The goal of any phase correction scheme is to recover, as far as is possible, the original spectrum.



II. Background

A. The ideal Fourier transform spectrometer

A simple Michelson interferometer is shown in Figure 1. Light leaving the source strikes the beam splitter where part of the amplitude is transmitted into one arm of the interferometer and part is reflected into the other arm.

Figure 1. A simple Michelson interferometer. The arm numbered 2 is $x/2$ longer than the one numbered 1; light in the second arm travels an extra path difference x .

These two rays travel to the end mirrors and return to the beam splitter where they recombine; the combined beams go off to the detector. This light is (more or less) a TEM electromagnetic wave, with Fourier components having the form $\vec{E}(\omega)e^{i(\vec{k}\cdot\vec{r})}$, where $|\vec{k}| = \omega$ (in vacuum) in units where $c = 1$. Thus if ω is in sec^{-1} , $|\vec{r}|$ will be a time; if ω is 2π (frequency in cm^{-1}), $|\vec{r}|$ is in cm. The phase of the electromagnetic wave advances as it moves through the interferometer; the phase is also modified by the beam splitter, which has amplitude transmission coefficient $t_b e^{i\phi_t}$ and amplitude reflection coefficient $r_b e^{i\phi_r}$; and by the end mirrors, which have amplitude reflection coefficients $r_1 e^{i\phi_1}$ and $r_2 e^{i\phi_2}$. At the detector, the total electric field amplitude is the sum of the amplitudes from the two interferometer arms;

$$\vec{E} = \vec{E}_1 + \vec{E}_2 \tag{1}$$

with

$$\vec{E}_1 = \vec{E}_s(\omega)t_b e^{i\phi_t} r_1 e^{i\phi_1} r_b e^{i\phi_r} e^{i\omega R} \tag{2}$$

$$\vec{E}_2 = \vec{E}_s(\omega) r_b e^{i\phi_r} r_2 e^{i\phi_2} t_b e^{i\phi_t} e^{i\omega R} e^{i\omega x} \quad (3)$$

where R is the total optical path length for both rays and x is the extra path difference for ray 2.

The ideal interferometer has identical arms with perfect mirrors ($r_1 e^{i\phi_1} = r_2 e^{i\phi_2} = -1$) and a perfect beam splitter (one which sends equal energy into both arms: $r_b = t_b = 1/\sqrt{2}$). The total intensity at the detector is

$$I(x) = \int_0^\infty d\omega |\vec{E}|^2 \quad (4)$$

$$= \frac{1}{2} \int_0^\infty d\omega S(\omega) [1 + \cos(\omega x)] \quad (5)$$

where $S(\omega) = |\vec{E}_s(\omega)|^2$ is the power spectrum. At zero path difference, the intensity at the detector is

$$I_0 = \int_0^\infty d\omega S(\omega) \quad (6)$$

At zero path difference all of the source intensity is directed to the detector; none returns to the source.² At large path differences the intensity at the detector is just half the zero path difference intensity

$$I_\infty = \frac{1}{2} \int_0^\infty d\omega S(\omega) \quad (7)$$

because as $x \rightarrow \infty$ the $\cos(\omega x)$ term averages to zero, i.e. it is more rapidly varying with frequency than $S(\omega)$.

The interferogram is the quantity $[I(x) - I_\infty]$; it is the cosine Fourier transform of the spectrum. It turns out to be convenient to factor out the integrated source intensity I_0 . We define a normalized spectral intensity $P(\omega)$ by

$$P(\omega) = S(\omega)/I_0 \quad (8)$$

so that, by equation (6), $\int_0^\infty d\omega P(\omega) = 1$. The normalized interferogram is

$$\gamma(x) = [I(x) - I_\infty]/I_0 \quad (9)$$

with these definitions, equation (5) becomes

$$\gamma(x) = \frac{1}{2} \int_0^{\infty} d\omega P(\omega) \cos(\omega x) \quad (10)$$

The spectrum can be recovered by the inverse cosine Fourier transform of $\gamma(x)$.

$$P(\omega) = \frac{4}{\pi} \int_0^{\infty} dx \gamma(x) \cos(\omega x) \quad (11)$$

There are no phase errors in this ideal case. Only the inverse cosine transform is calculated. Because $\gamma(x)$ is symmetric (or even) about $x = 0$, $\gamma(-x) = \gamma(x)$, the same $P(\omega)$ will be returned by an inverse complex Fourier transform:

$$P(\omega) = \frac{2}{\pi} \int_{-\infty}^{\infty} dx \gamma(x) e^{-i\omega x} \quad (12)$$

where the factor of 2 comes from our having extended the lower limit of the integral to $-\infty$. $P(\omega)$ remains a real function because the symmetry of $\gamma(x)$ about $x = 0$ causes the sine transform to be zero.

B. Phase errors

In a real interferometer, the assumptions made in the previous section do not in general hold. The beam splitter is not perfect; nor are the two arms of the interferometer perfectly matched. The beam splitter efficiency, ϵ_b , is defined to be:

$$\epsilon_b \equiv 4 R_b T_b \equiv 4 |r_b e^{i\phi} r|^2 |t_b e^{i\phi} t|^2 \quad (13)$$

The phase shifts on transmission and reflection do not play a role in the interferogram (assuming the beamsplitter to be homogeneous) because each beam was once transmitted and once reflected by the beam splitter. The beam splitter efficiency may be incorporated into the power spectrum by redefining it to be $S(\omega) \equiv \epsilon_b(\omega) |\tilde{E}_s(\omega)|^2$. The other factors which modify the spectrum, such as filter and window transmittances, detector spectral responsivity etc., may be included in exactly the same way by regarding $S(\omega)$ to be the instrumental power spectrum. (Note, however, that

those factors which modify the phase response of the system, such as the ac electrical phase shift of the detector and preamplifier of a rapid-scanning interferometer, cannot be accounted for in such a simple way. Indeed, these phase shifts must be corrected in the same way as other phase errors.)

If the two arms are not identical, then $\phi_1 \neq \phi_2$. (We will assume that $r_1 \approx r_2 \approx 1$; any small differences in the reflectance of the two mirrors will not affect the phase.) Let ϕ be the difference in phase between the two arms; i.e. let $\phi_2 = \phi_1 + \phi$. Then the total intensity at the detector is

$$I(x) = \frac{1}{4} \int_0^{\infty} d\omega |\vec{E}_s(\omega)|^2 \epsilon_b(\omega) r_1^2 \left[2e^{i\omega x} e^{i\phi(\omega)} + e^{-i\omega x} e^{-i\phi(\omega)} \right] \quad (14)$$

$$= \frac{1}{2} \int_0^{\infty} d\omega S(\omega) \left\{ 1 + \cos[\omega x + \phi(\omega)] \right\} \quad (15)$$

Notice that when $x = \infty$,

$$I_{\infty} = \frac{1}{2} \int_0^{\infty} d\omega S(\omega) \quad (16)$$

$$\equiv \frac{1}{2} I_0 \quad (17)$$

while when $x=0$,

$$I(0) = \frac{1}{2} \int_0^{\infty} d\omega S(\omega) \left[1 + \cos\phi(\omega) \right] \quad (18)$$

If $\phi(\omega) \neq 0$, then $I(0) < I_{\infty}$; all the source intensity is not sent to the detector. Thus, when phase errors are present the zero path difference intensity is less than twice the large path difference value. A signature of phase error is an interferogram whose maximum intensity is less than twice the intensity at very large path differences.

With the definitions (8) and (9), equation (15) becomes

$$\gamma(x) = \frac{1}{2} \int_0^{\infty} d\omega P(\omega) \cos[\omega x + \phi(\omega)] \quad (19)$$

Because $\gamma(x)$ is a real function, $P(\omega)e^{i\phi(\omega)}$ must be Hermi-

tion, i.e.

$$P(-\omega)e^{i\phi(\omega)} = P(\omega)e^{-i\phi(\omega)} \quad (20)$$

making equation (19) become

$$\gamma(x) = \frac{1}{4} \int_{-\infty}^{\infty} d\omega P(\omega)e^{i\phi(\omega)} e^{i\omega x} \quad (21)$$

The inverse Fourier transform of $\gamma(x)$ gives the spectrum and phase:

$$P(\omega)e^{i\phi(\omega)} = \frac{2}{\pi} \int_{-\infty}^{\infty} dx \gamma(x)e^{-i\omega x} \quad (22)$$

C. Phase correction

Suppose $\phi(\omega)$ is known. Then

$$P(\omega) = e^{-i\phi(\omega)} \frac{2}{\pi} \int_{-\infty}^{\infty} dx \gamma(x)e^{-i\omega x} \quad (23)$$

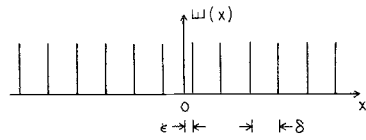
The spectrum $P(\omega)$ is found by multiplying the Fourier transform of the interferogram by $e^{-i\phi(\omega)}$. This procedure is the basis of the multiplicative technique first proposed by Mertz^{1b,3}. At least two alternative methods are used to eliminate the phase from equation (22). In one⁴, the power spectrum is calculated

$$|P(\omega)e^{i\phi(\omega)}| \equiv \sqrt{P_1(\omega)^2 + P_2(\omega)^2}$$

where $P_1(\omega) = P(\omega)\cos\phi(\omega)$ and $P_2(\omega) = P(\omega)\sin\phi(\omega)$ are the real and imaginary parts of the transform. This technique does correct phase errors but has a non-linear effect on noise and requires measurements for an equal distance on both sides of zero path difference.

The second alternative is to convolve the interferogram with the Fourier transform of $e^{-i\phi(\omega)}$ to produce a symmetric interferogram.⁵ The cosine inverse transform of this symmetric interferogram gives the spectrum. This convolution technique should be equivalent to the multiplicative technique; a common belief, however, is that the convolution

technique works well whilst the multiplicative one leads to baseline errors and other distortions in the spectrum.⁶ Even the fairly detailed studies by Sanderson and Bell⁸ of the multiplicative technique find residual errors around 0.5%, an unacceptably large value.



a)

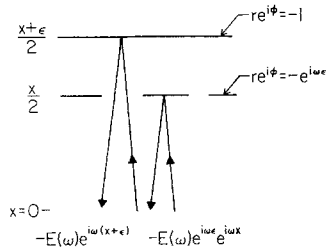
III. Sampling error is just another kind of phase error

In all experiments, the actual recorded interferogram is a sampled interferogram. The intensity is measured at points separated by uniform steps of length δ . Unless some care is taken, it is almost certain that the zero path difference position will be missed; the amount of the miss is $\epsilon \leq \delta/2$. Nevertheless, as we show below, so long as the spacing between sampled points is fine enough to avoid aliasing, the sampling procedure introduces no additional complications into phase correction.

The sampling procedure is represented mathematically by multiplying the interferogram by a "Dirac comb" or infinite sum of delta functions

$$\gamma_S(x) = \gamma(x) \sum_{k=-\infty}^{\infty} \delta(x - \epsilon - k\delta) \tag{24}$$

which picks out points at the sampling positions, as shown



b)

Figure 2a) The sampled interferogram is the product of the continuous interferogram and a set of equally spaced delta functions.

Figure 2b) A hypothetical mirror (dashed line) whose amplitude reflection coefficient is $re^{i\phi} = -e^{i\omega\epsilon}$ and which is located at $x/2$ is equivalent to a perfect mirror (solid line) with coefficient $re^{i\phi} = -1$ which is located at $(x+\epsilon)/2$.

in Figure 2a. The Fourier transform of this product is found by substituting γ_S for γ in equation (22).

$$P_S(\omega)e^{i\phi(\omega)} = \frac{2}{\pi} \int_{-\infty}^{\infty} dx \gamma(x) \sum_{k=-\infty}^{\infty} \delta(x-\epsilon-k\delta)e^{-i\omega x} \quad (25)$$

When the sum and the integral are interchanged, the integral may be done immediately, leaving the spectrum as the Fourier sum:

$$P_S(\omega)e^{i\phi(\omega)} = \frac{2}{\pi} \sum_{k=-\infty}^{\infty} \gamma(k\delta+\epsilon)e^{-i\omega k\delta} e^{-i\omega\epsilon} \quad (26)$$

or

$$P_S(\omega) = \frac{2}{\pi} e^{-i[\phi(\omega)+\omega\epsilon]} \sum_{k=-\infty}^{\infty} \gamma(k\delta+\epsilon)e^{-i\omega k\delta} \quad (27)$$

Equation (27) is an example of the shift theorem; shifting the origin of the Fourier transform by an amount ϵ introduces an overall phase factor $e^{-i\omega\epsilon}$ in the transform. Note that $P_S(\omega)$ remains continuous but is now periodic: $P_S(\omega+2\pi/\delta) = P_S(\omega)$.

Another way of viewing the errors introduced by missing the origin is to consider the interferometer mirror shown dashed in Figure 2b. Instead of being a perfect mirror at $(x+\epsilon)/2$, having $re^{i\phi} = -1$, it is a mirror at $x/2$ with $re^{i\phi} = -e^{i\omega\epsilon}$. Then the phase shift introduced by this mirror is $\phi = \omega\epsilon$, and its interferogram is given by equation (21) with $\phi(\omega) = \omega\epsilon$. The inverse Fourier transform of this interferogram is

$$P(\omega)e^{i\omega\epsilon} = \frac{2}{\pi} \int_{-\infty}^{\infty} dx \gamma(x)e^{-i\omega x} \quad (28)$$

This phase shift is just what the light traveling in that arm of the interferometer would have picked up on going on to the perfect mirror at $(x+\epsilon)/2$ and then returning to the beam splitter. Thus the interferometer with linear phase error $\omega\epsilon$ and the interferometer which has a systematic error ϵ in its path difference produce the same signal. The techniques used to correct phase errors in the continuous interferogram will also correct phase errors in the sampled interferogram.

IV. Phase correction and apodization

A. The effects of truncation

The values of path difference which an actual interferometer can produce are finite:

$$-L_1 \leq x \leq +L_2 \quad (29)$$

where we assume that one limit is on the negative side of zero path difference and the other is on the positive side and, for convenience, that $L_2 > L_1$. This truncation is equivalent to multiplication of the interferogram by an apodization-truncation function $A(x)$:

$$\gamma_A(x) = \gamma(x)A(x) \quad (30)$$

where

$$A(x) = \begin{cases} 0 & x < -L_1 \\ \text{finite} & -L_1 \leq x \leq L_2 \\ 0 & x > L_2 \end{cases}$$

The well-known^{1,9,10} effects of interferogram truncation are twofold: The instrumental resolution is limited to $\Delta\omega \approx 1/L_2$ and sidelobes or feet are introduced in the vicinity of sharp spectral features. The size of these feet may be reduced by wise choice of $A(x)$ over the range where it is finite, at the cost of further loss of resolution.

The Fourier transform of the product of any two functions is equal to the convolution of the Fourier transforms of the individual functions, so that

$$P_A(\omega) e^{i\phi_A(\omega)} = \frac{2}{\pi} \int_{-\infty}^{\infty} dx \gamma(x) A(x) e^{-i\omega x} \quad (32)$$

$$= [P(\omega) e^{i\phi(\omega)}] * a(\omega) \quad (33)$$

$$\equiv \int_{-\infty}^{\infty} d\omega' P(\omega') e^{i\phi(\omega')} a(\omega - \omega') \quad (34)$$

where

$$a(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dx A(x) e^{-i\omega x} \quad (35)$$

is the instrumental lineshape function.

It is this convolution which makes difficult the phase correction process. If $L_1 \neq L_2$ then $A(x)$ is not an even function of x and $a(\omega)$ is complex:

$$a(\omega) = a_1(\omega) + i a_2(\omega) \quad (36)$$

Because $a(\omega)$ is complex, $\phi_A(\omega)$ differs from the actual phase $\phi(\omega)$. Even in the case where the actual phase error is zero, off-center truncation causes a loss of contrast and a distortion of the baseline of the computed spectrum.

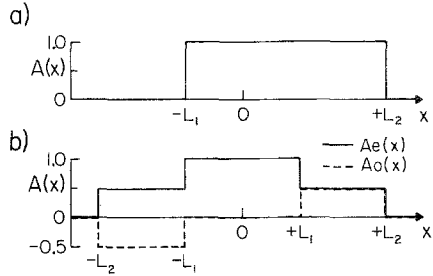


Figure 3a) An off-center boxcar function. The function $A(x)$ is unity over the range $-L_1, L_2$ and zero elsewhere.

Figure 3b) The even (solid line) and odd (dashed line) parts of the off-center boxcar function.

As an example, consider the off-center boxcar truncation function of Figure 3a. Here, $A(x) = 1$ over the interval $-L_1, L_2$. The even and odd parts of $A(x)$, given by

$$A_e(x) = \frac{1}{2} [A(x) + A(-x)] \quad (\text{even}) \quad (37)$$

$$A_o(x) = \frac{1}{2} [A(x) - A(-x)] \quad (\text{odd}),$$

are shown in Figure 3b. The real part of $a(\omega)$ is the Fourier transform of $A_e(x)$, which in this case is a superposition of two centered boxcar functions, one from $-L_1$ to $+L_2$ and the other from $-L_2$ to $+L_2$. This Fourier transform is

$$a_1(\omega) = \frac{1}{4\pi} \int_{-L_1}^{L_1} dx e^{-i\omega x} + \frac{1}{4\pi} \int_{-L_2}^{L_2} dx e^{-i\omega x} \quad (38)$$

$$a_1(\omega) = \frac{\sin(\omega L_1)}{2\pi\omega} + \frac{\sin(\omega L_2)}{2\pi\omega} \quad (39)$$

The convolution of $P(\omega)$ with $a_1(\omega)$ returns a spectrum that is the sum of a low resolution spectrum ($\Delta\omega \approx 1/L_1$) and a high resolution one ($\Delta\omega \approx 1/L_2$). The broad side lobes from the (mostly) unresolved features in the first spectrum cause baseline distortions; the lack of resolution itself causes loss of contrast.

A particularly sad case occurs when $L_1 \approx \delta$, the sampling interval of the interferogram. Now, the wings of the $\sin(\omega L_1)/2\pi\omega$ function extend above the aliasing frequency, $\omega_{\max} = 1/2\delta$, and are folded back into the spectrum, where they can pass through it several times.¹¹ As shown by Mertz³, this difficulty is alleviated by using an apodization function which does not double-weight the twice-measured portion of the interferogram between $-L_1$ and $+L_1$.

The imaginary part of the instrumental line shape, $a_2(\omega)$, can be neglected when $\phi(\omega) = 0$; the spectrum can be taken as the real part of $P_A(\omega)$. When $\phi(\omega) \neq 0$, however, both real and imaginary parts of $a(\omega)$ can contribute to the shape of the final spectrum.

B. Assumption of slowly varying phase.

We will assume in this section that the phase $\phi(\omega)$ is a slowly varying function of frequency. This assumption, if true, will enable us to determine the phase from a low-resolution two-sided interferogram, which extends from $-L_1$ to $+L_1$, and then use this phase to correct the high resolution interferogram, which extends to $+L_2$. This assumption is conventionally made in discussing phase correction procedures^{1, 3, 5-8}; if it is not true, a full two-sided interferogram would have to be measured.

We begin with equation (34) for the apodized spectrum and suppose that the phase $\phi(\omega')$ which appears in this equation varies slowly so that it can be expanded about frequency ω in a power series, and that only the first two terms need be kept

$$\phi(\omega') = \phi(\omega) + \left. \frac{d\phi}{d\omega'} \right|_{\omega'=\omega} (\omega' - \omega) \quad (40)$$

Note that we are not requiring that $\phi(\omega)$ have linear slope for all ω , only that it can be taken as linear about ω for the range of frequencies where $a(\omega - \omega') \neq 0$, i.e. for a few resolution widths. Then, with the definition

$$\beta(\omega) = \left. \frac{d\phi(\omega')}{d\omega'} \right|_{\omega'=\omega} \quad (41)$$

equation (34) becomes

$$P_A(\omega) e^{i\phi_A(\omega)} = \int_{-\infty}^{\infty} d\omega' P(\omega') e^{i[\phi(\omega) + \beta(\omega)(\omega' - \omega)]} a(\omega - \omega') \quad (42)$$

Notice that $e^{i\phi(\omega)}$ may be brought outside the integral and that $e^{i\beta(\omega)(\omega' - \omega)}$ has the same argument as the argument of $a(\omega - \omega')$ (except for sign). Thus, equation (42) is equivalent to

$$P_A(\omega) e^{i\phi_A(\omega)} = e^{i\phi(\omega)} [P(\omega)] * [a(\omega) e^{-i\beta\omega}] \quad (43)$$

$$= e^{i\phi(\omega)} [P(\omega)] * a^\beta(\omega) \quad (44)$$

The second term in the convolution is

$$\begin{aligned} a^\beta(\omega) &= a(\omega) e^{-i\beta\omega} = e^{-i\beta\omega} \frac{1}{2\pi} \int_{-\infty}^{\infty} A(x) e^{-i\omega x} dx \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} A(x - \beta) e^{-i\omega x} dx \end{aligned} \quad (45)$$

where the shift theorem has been invoked to move the origin (or center) of the apodization function to $x = \beta$. This result shows that the effects on the convolution of a linear phase error $\beta\omega$ may be eliminated by choosing the origin of the apodization function at $x = \beta$. If the phase error comes from a sampling error ϵ , then $\beta = \epsilon$ and the correct origin is the actual point of zero path difference.

Note that in going from equation (42) to equation (44) we have dropped the argument of β , so that the frequency-

dependent function $\beta(\omega)$ now appears to be a constant β . In principle, the frequency dependence remains in equations (44) and (45), apparently requiring a separate Fourier transform of the apodized interferogram $\gamma(x) [A(x-\beta(\omega))]$ at every frequency ω . Because by hypothesis $\beta(\omega)$ only varies over frequency intervals characteristic of the low resolution spectrum (i.e. $\Delta\omega \sim 1/L$) this requirement is even achievable in practice. Our test calculations show, however, that perfectly adequate results are obtained when β is taken as constant and $\phi(\omega)$ soaks up the remaining frequency dependence of the phase.

C. Phase correction of truncated and apodized interferogram.

For conceptual purposes, we consider apodization to be a two-step process. i.e.,

$$A(x) = A_I(x)A_{II}(x) \tag{46}$$

First, the central portion of the interferogram is multiplied by a ramp given by:

$$A_I(x) = \begin{cases} \frac{x+L_1}{2(\beta+L_1)} & -L_1 \leq x \leq L_1+2\beta \\ 1 & x \geq L_1+2\beta \\ 0 & \text{otherwise} \end{cases} \tag{47}$$

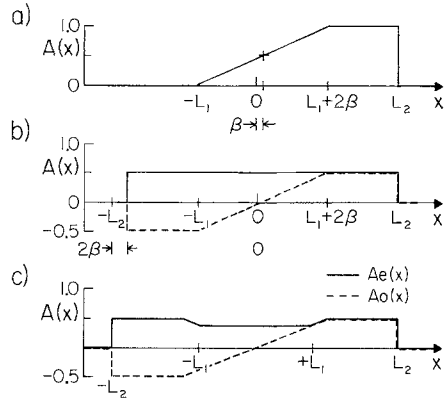


Figure 4a) The ramp through the origin gives proper weight to those interferogram points that were measured twice.

Figure 4b) The even (solid line) and odd (dashed line) parts of the left ramp.

Figure 4c) The even (solid line) and odd (dashed line) parts of a mislocated left ramp. The structure between $-L_1$ and $+L_1$ distorts the spectrum.

Note that $A_T(x) = \frac{1}{2}$ at $x=\beta$. The use of this ramp through the origin follows, essentially, the recommendation of Mertz and is shown in Figure 4a). Figure 4b) shows the even and odd parts (about $x=\beta$!) of this "left ramp". Note that the even part is a boxcar function which extends from $-L_2+2\beta$ to $+L_2$. Figure 4c) shows the even and odd parts of a left ramp which is not properly located. The structure near the origin in $A_{Te}(x)$ and $A_{To}(x)$ distorts the spectrum.

The second step in our apodization procedure is to multiply the data by the desired apodization function, $A_{II}(x)$. This function, whether triangular, Filler⁹, Norton-Beer¹⁰, or some other, is chosen so that its origin is also at $x=\beta$ and so that it is strictly an even function about this point, $A_{II}(-x-\beta) = A_{II}(x-\beta)$. For example, our triangular apodization function is

$$A_{II}(x) = \begin{cases} 1 - \frac{|x-\beta|}{L_2-\beta} & -L_2+2\beta \leq x \leq L_2 \\ 0 & \text{otherwise} \end{cases} \quad (48)$$

Note that $A_{II}(x) = 1$ at $x=\beta$ and goes to zero at $x = -L_2 + 2\beta$ and at $x=+L_2$. This choice of slope means that the periodic continuation of $A(x)$ has period $2(L_2-\beta)$ (which is twice the actual maximum optical path difference) and has a simple triangular waveform.

The inverse Fourier transform of the apodized interferogram can then be phase corrected. The phase corrected spectrum is

$$P_A(\omega) = \text{Re} \left\{ e^{-i\phi(\omega)} \frac{2}{\pi} \int_{-\infty}^{\infty} dx \gamma(x) A(x-\beta) e^{-i\omega x} \right\} \quad (49)$$

where the real part has been taken to eliminate the convolution of the spectrum with the imaginary part of the instrumental line shape, $a_2(\omega)$.

According to equation (44) this spectrum is

$$P_A(\omega) = P(\omega) * a_1^\beta(\omega) \quad (50)$$

To show explicitly that equation (49) does correct for

phase error, we consider the case of an intrinsically symmetric interferogram with a systematic error ϵ in the path difference (due, for example, to sampling errors). According to equation (28)

$$\phi(\omega) = \omega\epsilon$$

and $\beta = \epsilon$ at all frequencies. We use the convolution theorem on the right-hand side of equation 49 to find:

$$\begin{aligned} \operatorname{Re} \left\{ e^{-i\omega\epsilon} \frac{2}{\pi} \int_{-\infty}^{\infty} dx \gamma(x) A(x-\beta) e^{-i\omega x} \right\} &= \\ - \operatorname{Re} \left\{ e^{-i\omega\epsilon} \left[\frac{2}{\pi} \int_{-\infty}^{\infty} dx \gamma(x) e^{-i\omega x} \right] * \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} dx A(x-\beta) e^{-i\omega x} \right] \right\} & \\ = \operatorname{Re} \left\{ e^{-i\omega\epsilon} \left[p(\omega) e^{i\omega\epsilon} \right] * \left[a^{\beta}(\omega) \right] \right\} & \end{aligned}$$

If the instrumental line shape, $a^{\beta}(\omega)$, is sharp enough that $e^{i\omega\epsilon}$ is constant over the range of frequencies where $a^{\beta}(\omega - \omega') = 0$, then the term $e^{i\omega\epsilon}$ comes out of the convolution, and

$$P_A(\omega) = \operatorname{Re} \left\{ p(\omega) * a^{\beta}(\omega) \right\}$$

Because $P(\omega)$ is real, the real and imaginary parts of the instrumental lineshape do not mix in the convolution so that

$$P_A(\omega) = P(\omega) * a_1^{\beta}(\omega)$$

This result is the spectrum with high resolution, $\Delta\omega \approx 1/L_2$, determined by the width of $a_1^{\beta}(\omega)$.

V. Details of the procedure

A. Finding the phase

We determine $\phi(\omega)$ in the conventional way, by transforming a short interferogram which extends for nearly equal distances on either side of zero path difference. This short interferogram is picked out of the full inter-

ferogram $\gamma(x)$ by multiplying $\gamma(x)$ by $A_p(x)$ where

$$A_p(x) = \begin{cases} 0 & x < -L_1 \\ \text{finite} & -L_1 \leq x \leq L_1 + 2\beta' \\ 0 & x > L_1 + 2\beta' \end{cases} \quad (52)$$

As in the case of the apodization function for the high-resolution interferogram, $A_p(x)$ should be centered as closely as possible at zero path difference. We estimate this location by a parabolic fit to three points: the interferogram maximum and the points on either side of it. The parabola has a maximum at $x=\beta'$; we use this point as the center of our short interferogram. The Fourier transform of this short interferogram gives an amplitude and phase:

$$P_p(\omega)e^{i\phi(\omega)} = \frac{2}{\pi} \int_{-\infty}^{\infty} dx \gamma(x) A_p(x) e^{-i\omega x} \equiv P_{p1} + P_{p2} \quad (53)$$

with resolution $\sim 1/L_1$. The phase is calculated from

$$\phi(\omega) = \tan^{-1}(P_{p1}/P_{p2}) \quad (54)$$

If $P_{p1} < 0$, then π is subtracted from $\phi(\omega)$ when $P_{p2} < 0$ and π is added to $\phi(\omega)$ when $P_{p2} > 0$, giving a range of $-\pi$ to $+\pi$ for $\phi(\omega)$.

Note that because phase correction only requires $e^{-i\phi(\omega)}$ it is quicker to calculate

$$e^{-i\phi(\omega)} = (P_{p1} - iP_{p2}) / \sqrt{P_{p1}^2 + P_{p2}^2}$$

than to take the inverse tangent, put ϕ into the proper quadrant, and exponentiate. However, we also need $\phi(\omega)$ to estimate β , the coefficient of the linear term of the phase.

B. A note on the FFT

The fast Fourier transform (FFT) routine¹¹ calculates the inverse transform of an array γ_k by computing

$$P_j e^{i\phi_j} = \sum_{k=1}^M \gamma_k e^{-2\pi i k j / N} \quad (55)$$

for M (= a power of 2) values of j . In this equation we are labeling indices as in FORTRAN. Forman¹² has shown that in order to retrieve the phase, γ_1 should contain the first interferogram point for which $x \geq 0$. The other points from the positive x side follow in order. The remaining data (all of which has $x < 0$) are put into the top of the array. The last element, γ_M , contains the point that was originally adjacent to γ_1 . The elements between the last point with $x > 0$ and the first with $x < 0$ are set to zero. Flipping the interferogram about the origin in this way is tantamount to assuming a periodic interferogram, with period M ; the array contains the first period along positive x . Figure 5 shows how a schematic interferogram is set up for the FFT and phase retrieval.

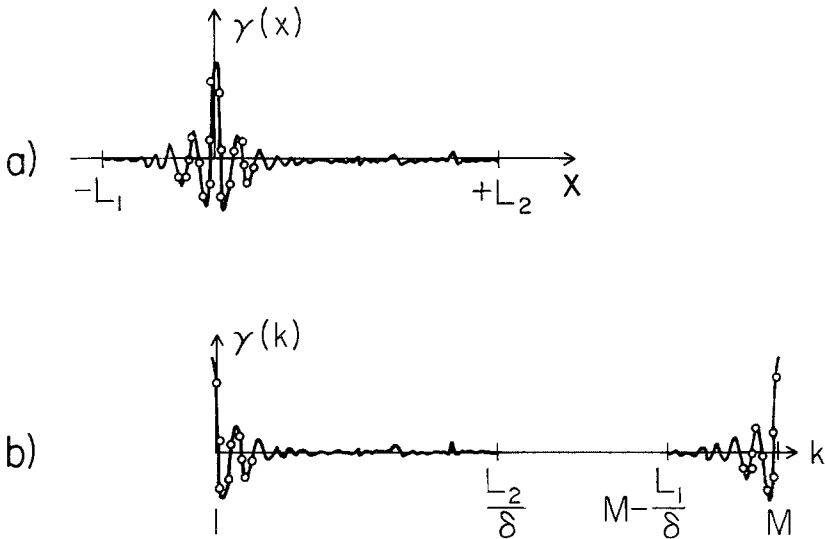


Figure 5a) Interferogram $\gamma(x)$ as recorded, extending from $-L_1$ to L_2 . The sampled points are shown near the origin.

Figure 5b) Interferogram $\gamma(k)$ set up for the FFT. There are M (a power of 2) points, with the array elements between $\frac{+L}{\delta}$ and $M - \frac{L_1}{\delta}$ being zero.

The number of points in the high resolution spectrum is $M \geq 2L_2/\delta$ (δ is the sampling interval), while the number in the low resolution spectrum (used to obtain the phase) is $M_p \geq 2L_1/\delta$. Because the phase is needed at every point of the high resolution spectrum, it is convenient to take $M_p = M$; this choice leads to a "zero-filled" interferogram and interpolated low-resolution spectrum.

C. Thirteen steps.

Our procedure for obtaining phase-corrected spectra is outlined in the following list:

1. Record a sampled interferogram from $x = -L_1$ to $x = +L_2$. L_1 should be large enough to resolve the broad features of the spectrum while L_2 gives the ultimate resolution.
2. Suppress the average value (I_ω) of the interferogram.
3. Search for the maximum interferogram point and fit a parabola through this point and the two adjacent ones. The peak of this parabola gives β' , the first estimate of the origin.
4. Pick out a two-sided interferogram with an equal number of points on either side of $x = \beta'$. This interferogram has $N_p = 2(L_1 + \beta')/\delta$ points.
5. If this interferogram is to be apodized, the apodization function should be centered at $x = \beta'$.
6. The minimum number of points where the phase is required is $N = 2(L_2 - \beta')/\delta$. Determine M , a power of 2 which is larger than N . Put the two-sided interferogram in an M -point array with the first point for which $x \geq \beta'$ as the first array element followed by the other points with $x > \beta'$. The remaining data points are put at the top of the array, with the last element being the last point for which $x < \beta'$. The center of this array contains many zeroes.
7. Compute the inverse complex Fourier transform.
8. Calculate the amplitude and phase spectra. Using a least squares fit weighted by the amplitude, find the constant β in $\phi(\omega) = \phi_0 + \beta\omega$.
9. Multiply the center section of the full interferogram by a ramp which is zero at $x = -L_1$, $\frac{1}{2}$ at $x = \beta$ and 1 at $x = L_1 + 2\beta$.
10. Apodize (about $\beta!$).

11. Set up an M-point array with the first element being the first data point with $x \geq \beta$ and continuing to $x = +L_2$. The array has zeroes where the data for $-L_2 \leq x < -L_1$ would have been. The data for $-L_1 \leq x < \beta$ are in the top of this array.

12. Compute the inverse Fourier transform, getting a complex array $P_1(\omega) + i P_2(\omega)$.

13. Calculate the phase corrected spectrum.

$$P(\omega) = \text{Re} \left\{ e^{-i\phi(\omega)} \left[P_1(\omega) + i P_2(\omega) \right] \right\}$$

Because the phase calculated in step 8 was interpolated, to M points, there will be a value for the phase at every point where the spectrum was calculated.

VI. Examples.

We have used the synthetic spectrum illustrated in Figure 6 to test our phase-correction procedures. This spectrum is supposed to represent a narrow emission peak at low frequencies and a broad continuum containing an absorption line at higher frequencies. The spectrum was synthesized at 1024

points, and reflected about $\omega = 0$ [$P(-\omega) = P(\omega)$] to give a 2048 point symmetric spectrum. This spectrum was phase shifted by multiplying by $e^{i\phi(\omega)}$, [$\phi(-\omega) = -\phi(\omega)$]. A complex FFT of this array gave a 2048-point interferogram which was real but asymmetric. A part of this interferogram, extending from $-L_1$ to $+L_2$, was the input interferogram for the phase correction process. The number of points in this test interferogram was always less than 1024, typically 512 on the positive side of the origin and a somewhat smaller number on the negative side.

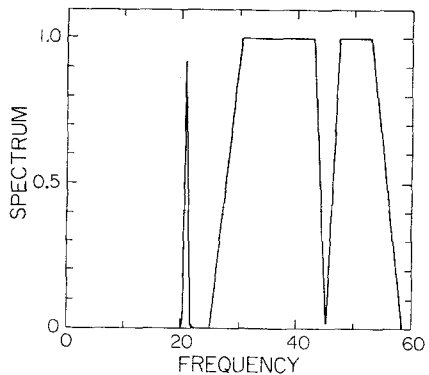


Figure 6. Synthetic spectrum for testing the phase-correction technique.

A. Proper phase correction

Figure 7 shows the amplitude and phase calculated from a 100 point two-sided interferogram which had a relatively large ($\epsilon = -0.98$) linear phase error. No apodization was used, so large sidelobes appear on the lower frequency line, which was not resolved. The phase has been suppressed at frequencies where the original spectrum had zero intensity.

The upper panel in Figure 8 gives the real and imaginary parts of the uncorrected spectrum, obtained from a 1024 point transform of $\gamma(x)$, where $50\delta \leq x \leq +512\delta$. The linear phase error, truncation, and ramp through the center all conspire to give a very disturbed spectrum. The lower panel shows the fully corrected spectrum. It is almost an exact replica of the original.

Figure 9 shows the residual error remaining after phase correction. (The residual error is defined here as the absolute value of the difference spectrum between the original spectrum and the

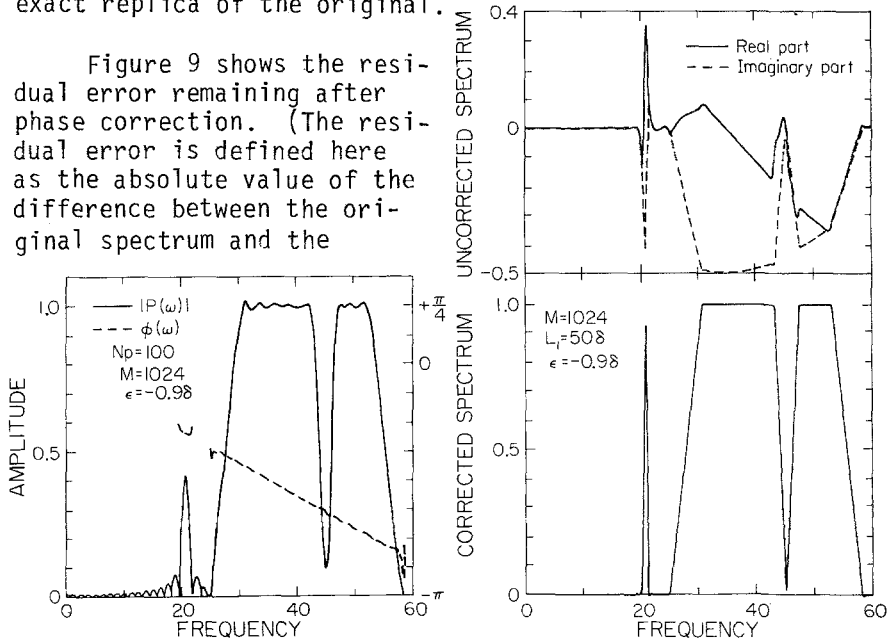


Figure 7. Low-resolution amplitude (solid line) and phase (dashed line) spectrum calculated from 100 points of an asymmetric interferogram and interpolated to 1024 output points.

Figure 8a) Real (solid line) and imaginary (dashed line) parts of the Fourier transform of an asymmetric interferogram that was truncated with the "left ramp" apodization function

Figure 8b) Phase corrected spectrum.

corrected spectrum and is expressed as a percentage of the spectrum maximum. Fine structure in the residual error has been smoothed over in making these plots.) The dashed line gives the error (about 1.5%) when the ramp is centered on the highest point of the interferogram (rather than $x+\beta$) while the solid line shows the error (about 0.02%) when the ramp is properly centered. The dotted line shows the error when the power spectrum $\sqrt{P_1^2 + P_2^2}$ is calculated from a complete double-sided interferogram. At many frequencies, the power spectrum has larger errors than the phase-corrected spectrum, suggesting that in both cases these residual errors result from round-off errors in the computer. Note these $\sim 0.02\%$ errors are substantially smaller than the $\sim 0.5\%$ errors reported by Sanderson and Bell.⁸

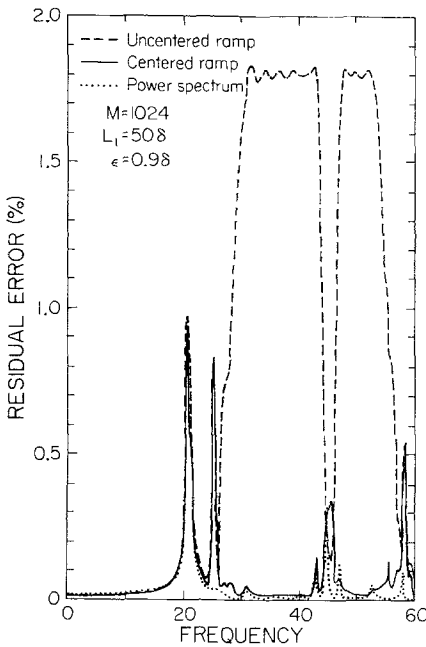


Figure 9. Residual error remaining after phase correction of an interferogram which had linear phase error. The solid line is found when the ramp is properly centered; the dashed line occurs when the ramp is improperly located; the dotted line is from a power spectrum.

For the incorrectly centered ramp, the residual errors scale with the magnitude of the phase error. In every case, however, the residual error is substantially

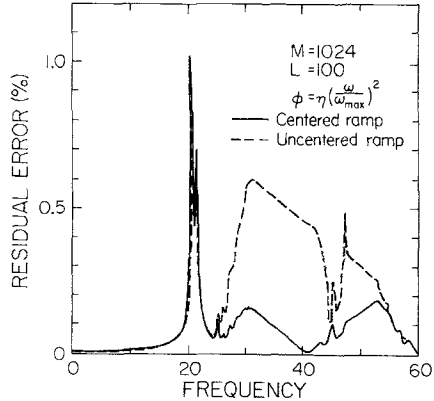


Figure 10. Residual error remaining after phase correction of an interferogram which had quadratic phase error. The solid line gives the error which occurs when the the ramp is properly centered while the dashed line gives the error when the ramp is improperly located.

reduced when the ramp is centered.

Figure 10 shows the residual error which occurs when there is a quadratic phase error, $\phi = \pi(\omega^2/\omega_{\max}^2)$. With a quadratic phase, the least squares fit to the phase produces a straight line through the curved quadratic phase error. Even though there is no well-defined origin of the interferogram when the phase error is quadratic, centering the truncation-apodization function at $x = \beta$ gives smaller residual error because the function is centered at the effective zero path difference ("point of stationary phase") of the most intense fringes.

B. Mistakes

In this section we show ways in which phase correction can go wrong. Throughout we use a modest value for the phase error, $\epsilon = 0.18$ rather than the value of 0.98 used previously. Nevertheless, these mistakes produce large errors. Figure 11 shows two examples of mis-weighting the interferogram center. The apodization functions used are shown at the bottom of the figure. The spectrum with the solid line was produced when the interferogram only included points $x > 0$, i.e., only a half-sided Fourier transform was done. This procedure produces the dreaded baseline tilt which has given the multiplicative technique a bad name.^{4,6-8}

For the dashed lines in Figure 11, all interferogram points were used without weighting at all; the apodization function is unity from $-L_1$ to $+L_2$. This function

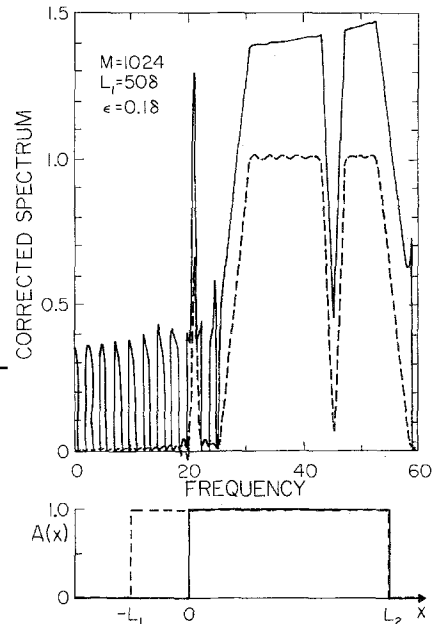


Figure 11a) Corrected spectrum (not residual error!) for truncation at $x=0$ (solid line) and for boxcar weighting of entire interferogram (dashed line).

Figure 11b) Truncation functions used to produce the spectra given above.

double weights the interferogram points between $-L_1$ and $+L_1$, producing the sum of a high-resolution spectrum and a low resolution spectrum. The result is a loss of intensity in the sharp features and extra structure near them.

Figure 12 shows in the solid line the result of using too few points for the two-sided interferogram. Here $L_1 = 12\delta$ as opposed to the value of 50δ used earlier. Severe distortions occur near sharp features in the spectrum because of poor resolution in the phase, even though the ramp was properly centered. (An anomaly occurs for the special case $\epsilon = 0.5$. For this value of phase error, the phase spectrum is very accurate; and a good corrected spectrum, shown as the dashed line in Figure 12, is obtained.

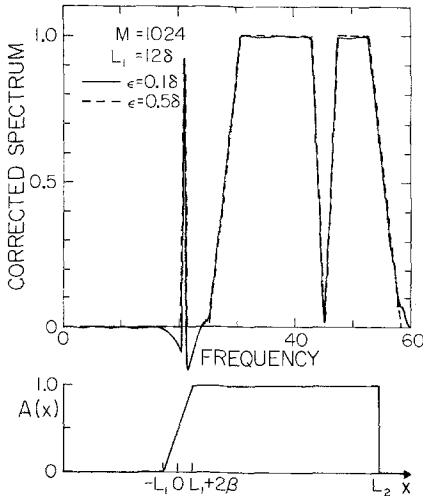


Figure 12a) The corrected spectrum obtained when $\epsilon=0.18$ and when only 12 points are taken before $x=0$ is shown as the solid line. The corrected spectrum obtained when $\epsilon = 0.58$ is shown as a dashed line.

Figure 12b) Truncation-apodization function used to obtain the above spectra.

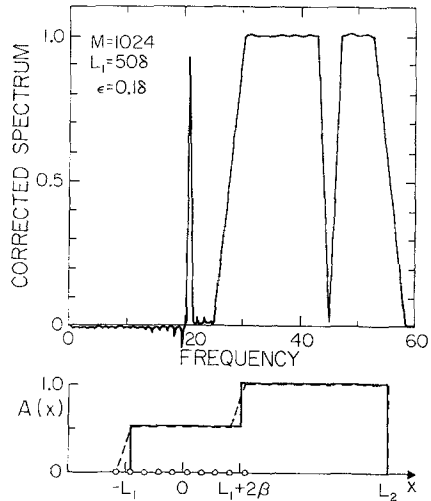


Figure 13a) The corrected spectrum found when a step is used as the truncation-apodization function.

Figure 13b) The solid line shows the truncation-apodization function used to obtain the above spectrum. The dashed line shows an equivalent truncation-apodization function when the interferogram is sampled at the values of x circled on the x -axis.

This anomaly occurs because the interferogram with $\epsilon = 0.5$ is symmetric about $x=0$, although no point corresponds to $x=0$.) In general, we have found that $L_1 \approx 50\delta$ produces good correction of linear phase error, with only small improvements if L_1 is increased. The correction of quadratic phase error seemed to require somewhat more points. There was a substantial improvement on going from $L_1 = 50\delta$ to $L_1 = 100\delta$, with only small improvements beyond this value.

One might think that multiplying the portion of the interferogram between $-L_1$ and $+L_1$ by one half should work as well as the ramp which we have been using. The result of this procedure is shown in Figure 13. The step weighting produces strong ripples in the spectrum. These ripples are the result of the attempt to make a step between two sampling points. As far as the interferogram is concerned, the apodization function contains two small ramps, as shown in the dashed line in the lower panel. These ramps are pinned to sampling points (shown as circles) and therefore cause structure to occur in the even part of the apodization function $A_e(x)$ at $x = \pm L_1$, which, when Fourier transformed and convolved with the spectrum, produce the ripples seen in Figure 13.

VII. Conclusions.

We have shown that a phase-corrected spectrum can be produced with an accuracy of $\pm 0.02\%$. To achieve this accuracy, it is essential that the truncation and apodization functions be properly located. The centers of the even parts of these functions must be at the best possible estimate of zero path difference.

References

1. Reviews of Fourier transform spectroscopy include:
 - a. G.A. Vanasse, A.T. Stair, Jr., and D.J. Baker, eds., Aspen International Conference on Fourier Spectroscopy, 1970.
 - b. L. Mertz, Transformation in Optics. (Wiley, New York, 1965).
 - c. R.J. Bell, Introductory Fourier Transform Spectroscopy. (Academic Press, New York, 1972).
 - d. R.B. Sanderson in Molecular Spectroscopy: Modern Research, K.N. Rao and C.W. Matthews, eds. (Academic Press, New York, 1972).
 - e. P.R. Griffiths, ed. Transform Techniques in Chemistry. (Plenum Press, New York, 1978).
 - f. A.E. Martin, Infrared Interferometric Spectrometers. (Elsevier, New York, 1980).
2. This point is not always recognized; see, for example pp. 203-205 of ref. 1c.
3. L. Mertz, Infrared Phys. 7, 17 (1967).
4. J. Connes, Rev. Opt. 40, 45 (1961).
5. M.L. Forman, W.H. Steel, and G.A. Vanasse, J. Opt. Soc. Am. 56, 59 (1966).
6. H. Sakai, G.A. Vanasse, and M.L. Forman, J. Opt. Soc. Am. 58, 84 (1968).
7. J. Connes in Aspen International Conference on Fourier Spectroscopy, 1970, G.A. Vanasse, A.T. Stair, Jr., and D.J. Baker, eds. (Air Force Cambridge Labs Report, No. 114), p. 83.
8. R.B. Sanderson and E.E. Bell, Applied Opt. 12, 266 (1973).
9. A.S. Filler, J. Opt. Soc. Am. 54, 762 (1964).

10. R.H. Norton and R. Beer, J. Opt. Soc. Am. 66, 259 (1976).
11. J.W. Cooley and J.W. Tukey, Math, Comput. 19, 297 (1965).
12. M.L. Forman, J. Opt. Soc. Am. 56, 978 (1966).