

Werner Herres and Joern Gronholz

Understanding FT-IR Data Processing

Part 1: Data Acquisition and Fourier Transformation

1 Introduction

Although infrared spectroscopy is one of the most powerful tools available to the analytical chemist and is routinely used in research and application labs and for process control, the most advanced form of IR-spectroscopy, Fourier Transform Infrared Spectroscopy (FT-IR), still holds some secrets for the chemist who is trained to work with conventional grating instruments. One reason is surely that the generation of the spectral trace is not straightforwardly controlled by setting appropriate knobs controlling slit widths, scanning speed, etc. but involves a certain amount of mathematical manipulations such as Fourier transformation, phase correction, and apodization, which may introduce a barrier to understanding the FTIR technique. Despite this difficulty, moderately and low priced FT-IR instruments are now entering even routine labs, because of their clear advantages compared to grating spectrometers. Even in lower priced FT-IR spectrometers, a laboratory, or dedicated computer is the most important component apart from the optics. As the quality of its software directly determines the accuracy of the spectra, it is recommended that the user be familiar with the principles of FT-IR data collection and manipulation. Unfortunately, there still seems to be a lack of literature on FT-IR at an introductory level. Therefore, this series of articles attempts to compile the essential facts in a, hopefully, lucid way without too many mathematical and technical details and thus provide an insight into the interrelation between FT-IR hardware, the data manipulations involved, and the final spectrum.

This is the first of a series of three articles, describing the data acquisition and mathematics performed by the minicomputer inside a FT-IR spectrometer. Special emphasis is placed on operations and artifacts relating to the Fourier transformation and on methods dealing directly with the Interferogram. Part 1 covers the measurement process and the conversion of the raw data (the Interferogram) into a spectrum.

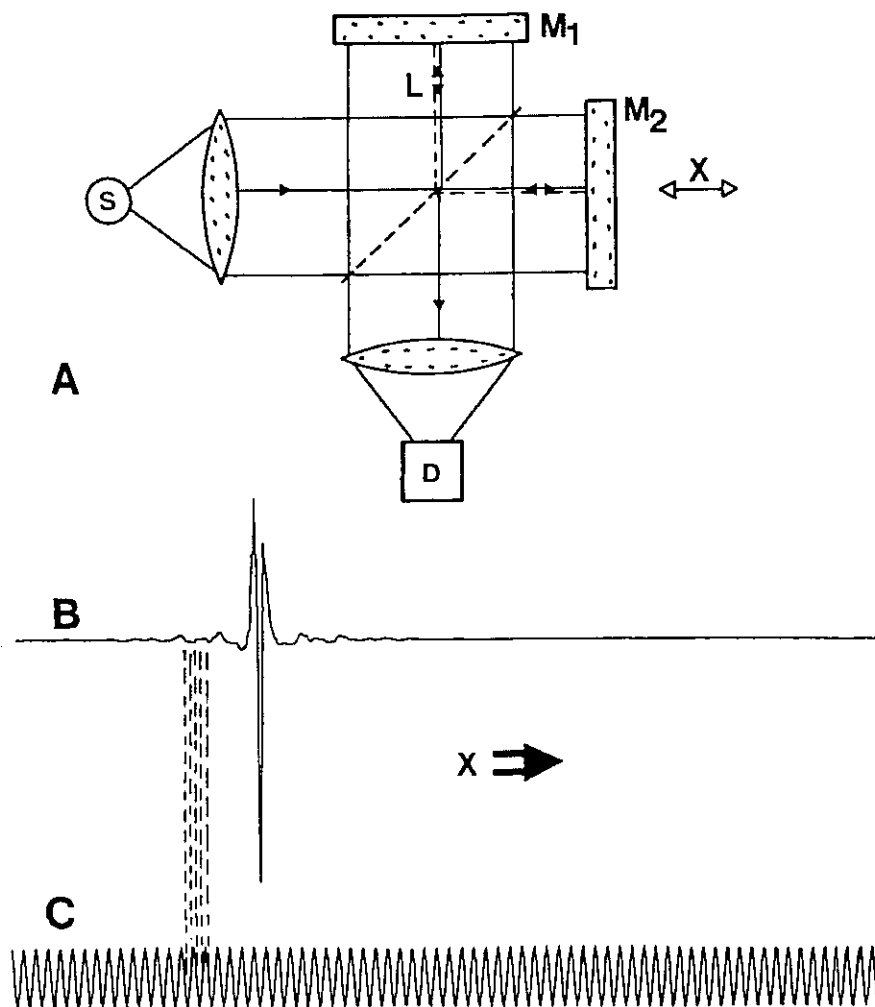


Figure 1: A) Schematics of a Michelson interferometer, S: source, D: detector, M1: fixed mirror, M2: movable mirror, X: mirror displacement, B) Signal measured by detector D. This is the Interferogram. C) Interference pattern of a laser source. Its zero crossings define the positions where the interferogram is sampled (dashed lines).

2 Raw Data Generation

The essential piece of optical hardware in a FT-IR spectrometer is the interferometer. The basic scheme of an idealized Michelson interferometer is shown in Figure 1.

Infrared light emitted by a source (Globar, metal wire, Nernst bar...) is directed to a device called the beam splitter, because it ideally allows half of the light to pass through while it reflects the other half.

The reflected part of the beam travels to the fixed mirror M1 through a distance L , is reflected there and hits the beam splitter again after a total path length of $2L$. The same happens to the transmitted part of the beam. However, as the reflecting mirror M2 for this interferometer arm is not fixed at the same position L but can be moved very precisely back and forth around L by a distance x , the total path length of this beam is accordingly $2*(L+x)$. Thus when the two halves of the beam recombine again on the beam splitter they exhibit a path length difference or optical retardation of $2*x$, i.e. the partial beams are spatially coherent and will interfere when they recombine.

The beam leaving the interferometer is passed through the sample compartment and is finally focused on the detector D. The quantity actually measured by the detector is thus the intensity $I(x)$ of the combined IR beams as a function of the moving mirror displacement x , the so-called interferogram (Figure 1 B).

The interference pattern as seen by the detector is shown in Figure 2A for the case of a single, sharp spectral line. The interferometer produces and recombines two wave trains with a relative phase difference, depending on the mirror displacement. These partial waves interfere constructively, yielding maximum detector signal, if their optical retardation is an exact multiple of the wavelength λ , i.e. if

$$2*x = n*\lambda \quad (n=0,1,2,\dots) \quad (1)$$

Minimum detector signal and destructive interference occur if $2*x$ is an odd multiple of $\lambda/2$. The complete dependence of $I(x)$ on x is given by a cosine function:

$$I(x) = S(\nu) * \cos(2\pi\nu x) \quad (2)$$

Where we have introduced the wavenumber $\nu = 1/\lambda$, which is more common in FT-IR spectroscopy, and $S(\nu)$ is the intensity of the monochromatic line located at wavenumber ν .

Equation (2) is extremely useful for practical measurements, because it allows very precise tracking of the movable mirror. In fact, all modern FT-IR spectrometers use the interference pattern of the monochromatic light of a He-Ne laser to control the change in optical path difference. This is the reason why we included the interference pattern of the He-Ne laser in Figure 1C. This demonstrates how the IR interferogram is digitized precisely at the zero crossings of the laser interferogram. The accuracy of the sample spacing Δx between two zero crossings is solely determined by the precision of the laser wavelength itself. As the sample spacing $\Delta \nu$ in the spectrum is inversely proportional to Δx , the error in $\Delta \nu$ is of the same order as in Δx . Thus, FT-IR spectrometers have a built-in wavenumber calibration of high precision (practically about 0.01 cm^{-1}). This advantage is known as the Connes advantage.

3 Advantages of FT-IR

Besides its high wavenumber accuracy, FT-IR has other features, which make it superior to conventional IR.

The so-called Jacquinot or throughput advantage arises from the fact that the circular apertures used in FT-IR spectrometers have a larger area than the linear slits used in grating spectrometers, thus enabling higher throughput of radiation.

In conventional spectrometers the spectrum $S(\nu)$ is measured directly by recording the intensity at different monochromator settings ν , one ν after the other. In FT-IR, all frequencies emanating from the IR source impinge simultaneously on the detector. This accounts for the so-called multiplex or Fellgett advantage.

The measuring time in FT-IR is the time needed to move mirror M2 over a distance proportional to the desired resolution. As the mirror can be moved very fast, complete spectra can be measured in fractions of a second. This is essential, e.g. in the coupling of FT-IR to capillary GC, where a time resolution of 10 - 20 spectra per sec-

ond at a resolution of 8 cm^{-1} is often necessary [1].

Finally, the Fellgett and Jacquinot advantages permit construction of interferometers having much higher resolving power than dispersive instruments.

Further advantages can be found in the IR literature, e.g. in the book by Bell [2].

4 Fourier Transformation

Data acquisition yields the digitized interferogram $I(x)$, which must be converted into a spectrum by means of a mathematical operation called Fourier transformation (FT). Generally, the FT determines the frequency components making up a continuous waveform. However, if the waveform (the interferogram) is sampled and consists of N discrete, equidistant points, one has to use the discrete version of the FT, i.e. discrete FT (DFT):

$$S(k\Delta\nu) = \sum_{n=0}^{N-1} I(n\Delta x) \exp(i2\pi nk/N) \quad (3)$$

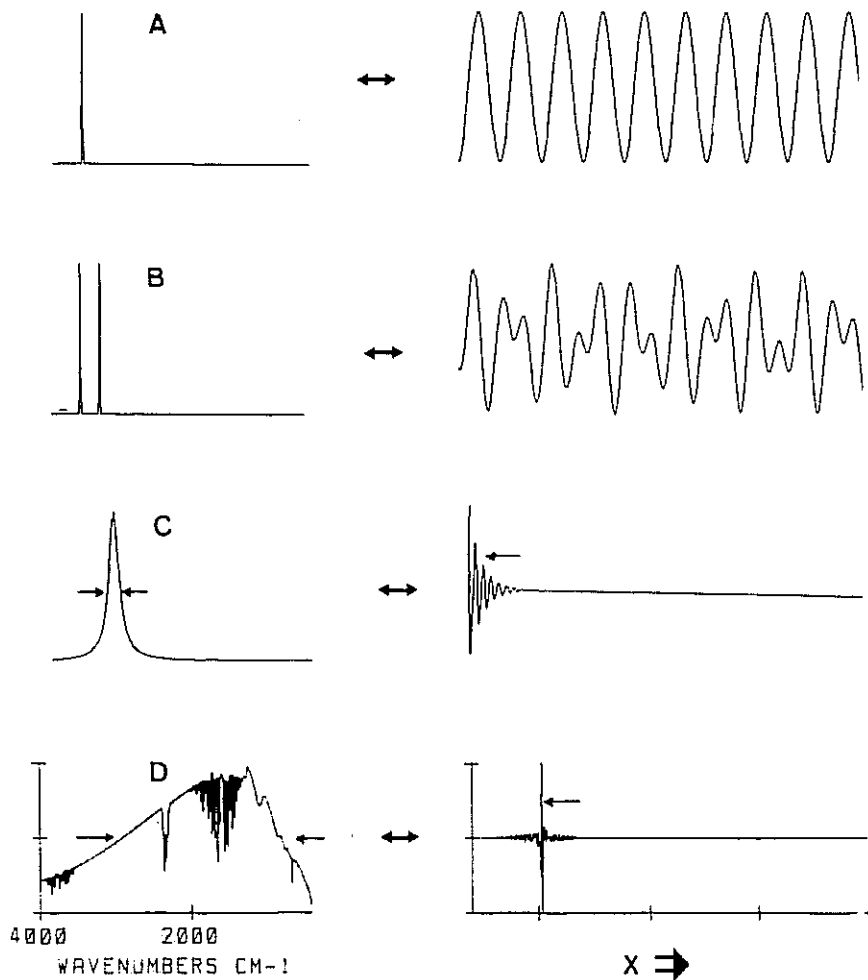
Where the continuous variables x , ν have been replaced by $n*\Delta x$ and $k*\Delta \nu$, respectively. The spacing $\Delta \nu$ in the spectrum is related to Δx by

$$\Delta \nu = 1/(N * \Delta x) \quad (4)$$

The DFT expresses a given function as a sum of sine and cosine functions. The resulting new function $S(k*\Delta \nu)$ then consists of the coefficients (called the Fourier coefficients) necessary for such a development. Alternatively, if the set $S(k*\Delta \nu)$ of Fourier coefficients is known, one can easily reconstruct the interferogram $I(n*\Delta x)$ by combining all cosines and sines multiplied by their Fourier coefficients $S(k*\Delta \nu)$ and dividing the whole sum by the number of points N . This is stated by the formula for the inverse DFT (IDFT):

$$I(n*\Delta x) = (1/N) \sum_{k=0}^{N-1} S(k*\Delta \nu) \exp(-i2\pi * nk/N) \quad (5)$$

The summation (5) is best illustrated in the simple case of a spectrum with one or two monochromatic lines, as shown in Figures 2A and 2B. For a limited number of functions like the Lorentzian in Figure 2C, the corresponding FT is known analytically



and can be looked up from an integral table. However, in the general case of measured data, the DFT and IDFT must be calculated numerically by a computer.

Although the precise shape of a spectrum cannot be determined from the interferogram without a computer, it may nevertheless be helpful to know two simple trading rules for an approximate description of the correspondence between $I(n \cdot \Delta x)$ and $S(k \cdot \Delta \nu)$.

From Figure 2C we can, e.g., extract the general qualitative rule that a finite spectral line width (as is always present for real samples) is due to damping in the interferogram: The broader the line the stronger the damping.

Comparing the widths at half height (WHH) of $I(n \cdot \Delta x)$ and $S(k \cdot \Delta \nu)$, reveals

another related rule: The WHH's of a 'hump-like' function and its FT are inversely proportional. This rule explains why in Figure 2D the interferogram due to a broad band source shows a very sharp peak around the zero path difference position $x=0$, while the wings of the interferogram, which contain most of the useful spectral information, have a very low amplitude. This illustrates the need for ADC's of high dynamic range in FT-IR measurements. Typically, FT-IR spectrometers are equipped with 15- or 16-bit ADC's.

For $n=0$, the exponential in (5) is equal to unity. For this case, expression (5) states, that the intensity $I(0)$ measured at the interferogram centerburst is equal to the sum over all N spectral intensities divided by N . This means the height of

the center burst is a measure of the average spectral intensity.

In practice, eq. (3) is seldom used directly because it is highly redundant. Instead a number of so-called fast Fourier transforms (FTT's) are in use, the most common of which is the Cooley-Tukey algorithm. The aim of these FTT's is to reduce the number of complex multiplications and sine- and cosine calculations appreciably, leading to a substantial saving of computer time. The (small) price paid for the speed is that the number of interferogram points N cannot be chosen at will, but depends on the algorithm. In the case of the Cooley-Tukey algorithm, which is used by most FT-IR manufacturers with slight modifications, N must be a power of two. For this reason and from relation (4) it follows that spectra taken with laser-controlled FT-IR spectrometers will show a sample spacing of $\Delta \nu = m \cdot \text{laser wavenumber} / 2 \cdot N$.

5 Final Transmittance Spectrum

To obtain a transmittance spectrum, the three steps shown in Figures 3 A, B, C are necessary (this example was taken from a GC run):

- An interferogram measured without sample in the optical path is Fourier transformed and yields the so-called single channel reference spectrum $R(\nu)$ of Fig. 3A.
- An interferogram with a sample in the optical path is measured and Fourier transformed. This yields the so-called single channel sample spectrum $S(\nu)$ of Fig. 3B. $S(\nu)$ looks similar to $R(\nu)$ but has less intensity at those wavenumbers where the sample absorbs.
- The final transmittance spectrum $T(\nu)$ is defined as the ratio $T(\nu) = S(\nu) / R(\nu)$. This is shown in Fig. 3C.

Once the transmittance spectrum has been obtained, further data processing resembles that of digitized spectra from dispersive instruments.

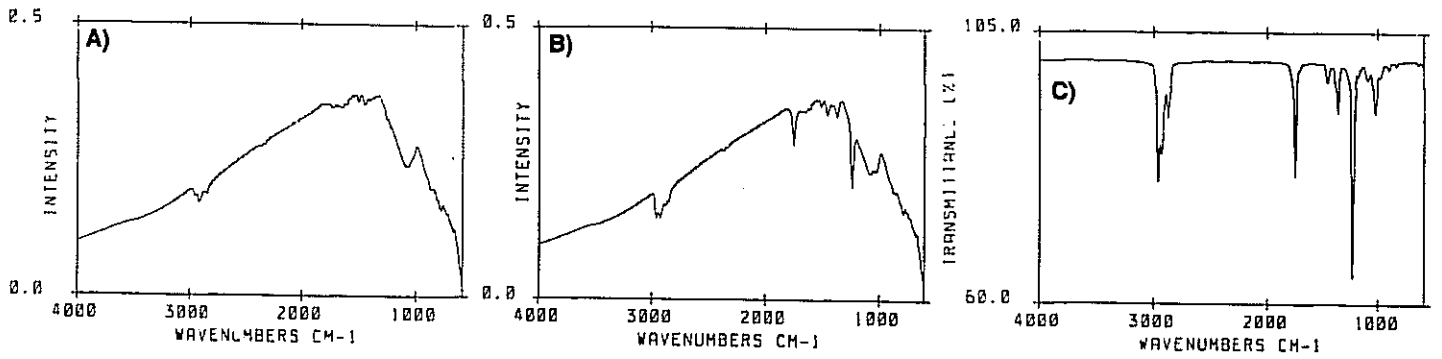


Figure 3: A) Single channel reference spectrum measured through an empty sample compartment. B) Single channel spectrum of absorbing sample. C) Transmittance spectrum equal to Fig. 3B divided by Fig. 3A.

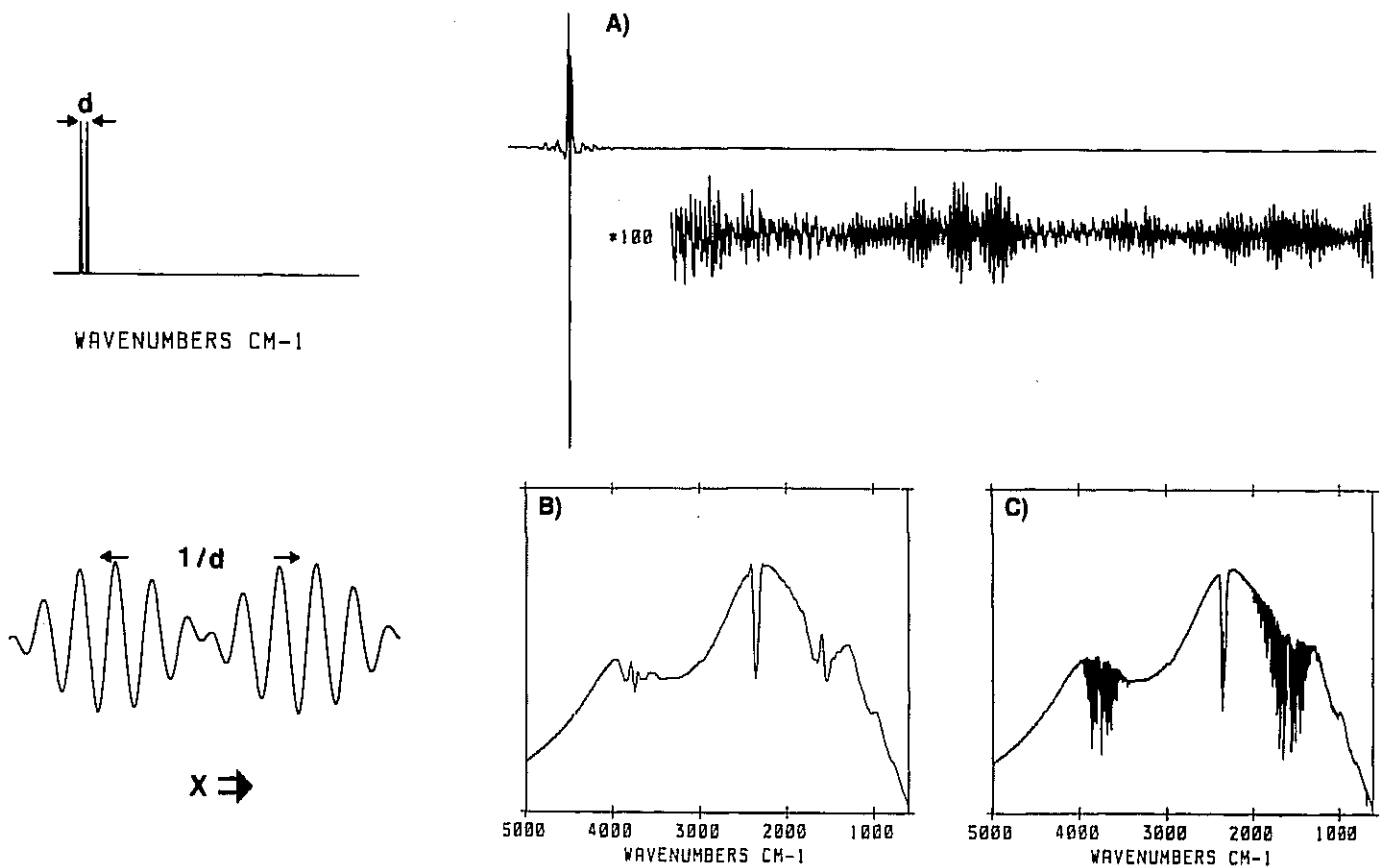


Figure 4: Two closely spaced spectral lines at distance d (left) produce repetitive patterns at distance $1/d$ in the interferogram (right).

Figure 5: A) First 2048 points of an Interferogram consisting of a total of 8196 points. Signal in the wings is amplified 100 times. B) FT of first 512 points of interferogram in Fig. 5a, corresponding to a resolution of 32 cm. C) FT of all 8196 points of interferogram in Fig. 5a, corresponding to a resolution of 2 cm.

6 Resolution in FT-IR

Figure 4 shows the interferogram corresponding to two sharp lines separated by a wavenumber distance d . Due to the separation d in the spectrum, the interferogram shows periodic modulation patterns repeated after a path length difference $1/d$. The closer the spectral lines are, the greater the distance between the repeated patterns. This illustrates the so-called Rayleigh criterion, which states that in order to resolve two spectral lines separated by a distance d one has to measure the interferogram up to a path length of at least $1/d$.

For a practical measurement, which was done on a Bruker IFS-88 using a broad band MCT detector, the influence of increasing the interferogram path length on the resolution is shown in Figures 5A, B, C. The interferogram in Figure 5A represents the first 2048 points from a total of 8196. Figure 5B was obtained by transforming only the first 512 interferogram points, which corresponds to a resolution of 32 cm^{-1} . Figure 5C exhibits the full 8196 point transform. It is clearly seen that many more spectral features are resolved in the case of a longer optical path.

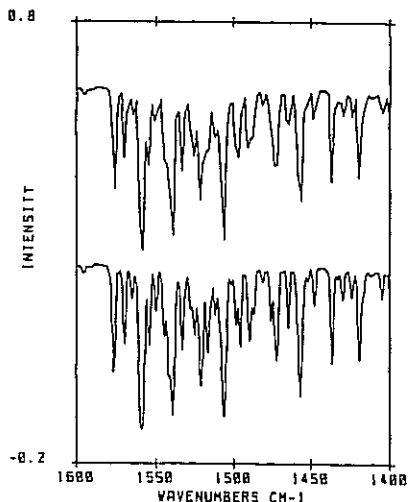
7 Zero-Filling

It should be noted that DIFT only approximates the continuous FT, although it is a very good approximation if used with care. Blind use of eq. (3), however, can lead to three well-known spectral artifacts: the picket-fence effect, aliasing, and leakage.

The picket-fence effect becomes evident when the interferogram contains frequencies which do not coincide with the frequency sample points $k \cdot \Delta\nu$. If, in the worst case, a frequency component lies exactly halfway between two sample points, an erroneous signal reduction by 36 % can occur: one seems to be viewing the true spectrum through a picket-fence, thereby clipping those spectral contributions lying 'behind the pickets', *i.e.* between the sampling positions $k \cdot \Delta\nu$. In practice, the problem is less extreme than stated above if the spectral components are broad enough to be spread over several sampling positions.

The picket-fence effect can be overcome by adding zeros to the end of the interferogram before DFT is performed,

thereby increasing the number of points per wave number in the spectrum. Thus, zero filling the interferogram has the effect of interpolating the spectrum, reducing the



References

- [1] W. Herres, "Capillary GC-FTIR Analysis of Volatiles: HRGC-FTIR" in Proceedings "Analysis of Volatiles", P. Schreier (editor), Walter de Gruyter & Co., Berlin (1984).
- [2] R. J. Bell, "Introductory Fourier Transform Spectroscopy", Academic Press, New York (1972).
- [3] G.D. Bergland, *IEEE Spectrum*, Vol. 6, pp. 41 - 52 (1969).

error. As a rule of thumb, one should always at least double the original interferogram size for practical measurements by zero filling it, *i.e.* one should choose a zero filling factor (ZFF) of two. In those cases, however, where the expected line width is similar to the spectral sample spacing (as e.g. in case of gas-phase spectra), a ZFF value of up to 8 may be appropriate.

The influence of zero filling on the appearance of water vapor bands is demonstrated in Figure 6. At the top, a spectrum with no zero filling is shown. The spectrum at the bottom is zero filled using a ZFF of 8. While the lines of the upper spectrum look badly clipped, the lines are smooth in zero filled spectrum.

It should be noted that zero filling does not introduce any errors because the instrumental line shape is not changed. It is therefore superior to polynomial interpolation procedures working in the spectral domain.

Aliasing, leakage, apodization, and phase correction will be dealt with in the following installments.

Series

Werner Herres and Joern Gronholz

Understanding FT-IR Data Processing

Part 2: Details of the spectrum calculation

In the first part of this series, we covered the FT-IR data acquisition and the Fourier transformation. This second part continues with the description of the mathematical operations performed by a FT-IR minicomputer to compute the spectrum from the interferogram.

1 Aliasing

In part 1 [1] it was shown that sampling the continuous interferogram and the use of the discrete version of the Fourier transformation (the DFT) might produce artifacts, such as the picket-fence effect, unless special precautions are taken. Another possible source of error due to the use of the DFT is aliasing.

To understand aliasing, we recall the basic DFT-equation

$$S(k * \Delta\nu) = \sum_{n=0}^{N-1} \exp(i2\pi kn/N) I(n * \Delta x) \quad (1)$$

which describes how a spectrum sampled at wavenumbers $k * \Delta\nu$ can be computed from an interferogram sampled at optical path differences $n * \Delta x$. In practical calculations both n and k will run from 0 to $N-1$, i.e. the DFT produces N (generally complex) output points from an input interferogram of N (generally real) input points. If we expect the spectrum to be of the form shown in Figure 1A, we will find that the DFT yields not just a single spectrum but rather the spectrum plus its mirror image, as given in Figure 1B: the first $N/2$ points represent the expected spectrum, the second part, starting with the index $k = N/2$ equals its mirror image. For practical computations this means that a DFT of an N -point interferogram yields only $N/2$ meaningful output points. The second set of $N/2$ points is redundant and therefore automatically discarded. This behavior is also easily derived from Equation (1), if

one substitutes index k by $N-1$. Using the identity

$$\begin{aligned} \exp i2\pi k &= (\exp i2\pi) ** k \\ &= 1 ** k \\ &= 1 \end{aligned} \quad (2)$$

one obtains the mathematical description of mirror symmetry

$$S([N - k]) = S(-k) \quad (3)$$

about the so-called 'folding'- or 'Nyquist'-wavenumber ν_f

$$\begin{aligned} \nu_f &= (n/2) * \Delta\nu \\ &= 1/(2 * \Delta x) \end{aligned} \quad (4)$$

Furthermore, one sees that Equation (1) is not only valid for indices k from 0 to $N-1$ but for all integers including negative numbers. In particular, if we replace k in Equation (1) by $k + m * N$, we get the equation

$$S([k + m * N]) = S(k) \quad (5)$$

which states that the mirror-symmetrical N -point sequence of Figure 1B is endlessly and periodically replicated as indicated in Figure 1C.

This replication of the origin spectrum and its mirror image on the wavenumber axis is termed 'aliasing'.

1.1 Alias Overlap

From Figures 1B and 1C it is clear that a unique spectrum can only be calculated if the spectrum does not overlap with its

mirror-symmetrical replicate (alias). No overlap will occur if the spectrum is zero above a maximum wavenumber ν_{\max} and if ν_{\max} is smaller than the folding wavenumber ν_f :

$$\begin{aligned} \nu_{\max} \leq \nu_f &= (N/2) * \Delta\nu \\ &= 1/(2 * \Delta x) \end{aligned} \quad (6)$$

Here, we recall from [1] that $\Delta\nu$ is related to Δx by

$$\Delta\nu = 1/(N * \Delta x) \quad (7)$$

If, however, the spectrum contains a non-zero contribution e.g. 200 cm^{-1} above the folding wavenumber ν_f , this will be 'folded back' below ν_f and appear at the wrong position, i.e. $\nu_f - 200 \text{ cm}^{-1}$. This is the possible artifact due to aliasing.

The finer the interferogram sample spacing Δx is, the further apart are the aliases and the lower the danger of alias overlap. However, small Δx means also an increased number of points N and therefore higher storage needs and larger computation times. For a given wavenumber range, the FT-IR software has therefore to choose the maximum sample spacing for which still no overlap occurs.

In part 1, we explained that in FT-IR, the sampling positions are derived from the zero crossings of an He-Ne laser wave having a wavelength λ of $1/15800 \text{ cm}$. As a zero crossing occurs every $\lambda/2$, the minimum possible sample spacing Δx_{\min} is $1/31600 \text{ cm}$. With Equation (4), this

corresponds to a folding wavenumber of 15800 cm^{-1} , i.e. the maximum bandwidth which can be measured without overlap has a width of 15800 cm^{-1} . A larger range can be covered, if the laser frequency is electronically doubled (frequency multiplication).

Very often, the investigated bandwidth is much smaller than 15800 cm^{-1} , e.g. in mid-IR, where ν_{\max} is generally less than 4500 cm^{-1} and especially in the far-IR with wavenumbers below 200 cm^{-1} . In these cases, one can choose Δx to be an m -fold multiple of Δx_{\min} . This leads to an m -fold reduction of the interferogram size.

1.2 Undersampling

An even greater reduction of the data size is possible, if the spectrum is zero below a lower band limit ν_{\min} and if ν_{\min} is not zero as assumed above. If the spectrum band limits ν_{\min} and ν_{\max} lie between lower and upper folding wavenumbers ν_{fL} and ν_{fU} which are related by

$$\nu_{fL} = \frac{n-1}{n} \nu_{fU} \quad (n = 1, 2, 3, \dots) \quad (8)$$

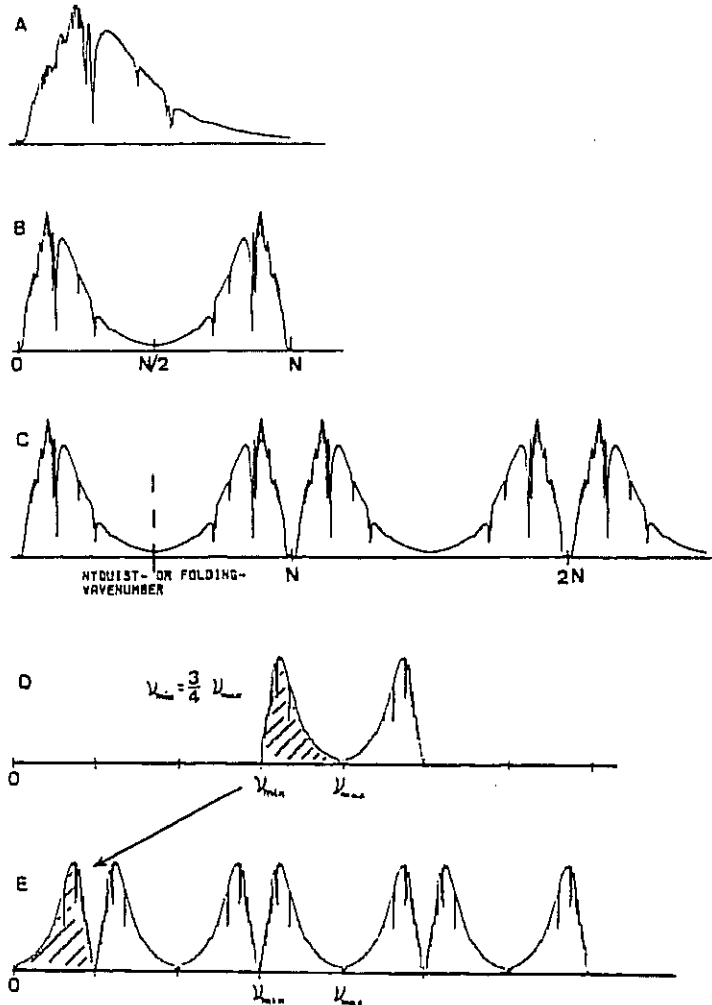
it will look as indicated in Figure 1D for the case $n = 4$. The upper folding wavenumber ν_{fU} must be a natural fraction (or integer multiple) of the He-Ne laser wavenumber:

$$\nu_{fU} = f \cdot 15800 \quad f = \dots, \frac{1}{3}, \frac{1}{2}, 1, 2, 3, \dots \quad (9)$$

If we now further increase the sample spacing by a factor n , the aliases of Figure 1D will overlap appreciably, thus filling the previously empty range from 0 to ν_{fL} with $n-1$ further copies of the original spectrum. This is shown in Figure 1E. As all copies are identical (except that their absolute wavenumber scaling and their direction on the ν -axis can differ from the

original), we need not calculate the spectrum at its true position by an N/n -point FT, but rather calculate the alias of lowest wavenumber by an N/n -point FT and correct its wavenumber scaling afterwards. This further n -fold reduction of interferogram size compared to the conventional case, where $\nu_{\min} = 0$, is

termed 'undersampling'. It should be noted that undersampling enables measurements with a ν_{\max} higher than the original laser wavenumber, because only the difference $\nu_{fU} - \nu_{fL}$ and not the absolute values of the folding wavenumbers is to be considered in the sampling condition



$$\Delta x \leq \frac{1}{2|v_{fU} - v_{fL}|} \quad (10)$$

An advanced FT-IR software package will automatically account for proper sampling and undersampling, such that only the upper and lower limits of the desired spectral range need to be specified. The user only needs to make sure that the investigated spectrum is really zero outside the range v_{fL} to v_{fU} by inserting

either optical or electronic filters.

For technical reasons, the sample spacing must often be increased in steps of powers of two. The possible folding wavenumbers for this common case are given in Table 1.

2 Effect of the Finite Record Length: Leakage

Unlike the picket-fence effect and aliasing,

leakage is not due to using a digitized version of a continuous interferogram. Leakage is caused by the truncation of the interferogram at finite optical path difference. The proper mathematical term to describe the effect on the spectra of truncating the interferogram is 'convolution'.

2.1 Convolution

Mathematically, an interferogram $IL(x)$, truncated at optical path difference $x = L$ can be obtained by multiplying an interferogram $Ii(x)$ of infinite extension by a 'boxcar' function $BX(x)$, which is zero for $x > L$ and unity for $x \leq L$, i. e.

$$IL(x) = Ii(x) * BX(x) \quad (11)$$

According to the convolution theorem of Fourier analysis, the Fourier transform of a product of two functions is given by the convolution (here indicated by the 'O' symbol of their individual Fourier transforms, i.e. if $Si(\nu)$ and $bx(\nu)$ are the Fourier transforms of $Ii(x)$ and $BX(x)$, respectively:

$$\begin{aligned} Si(\nu) &= \int_{-\infty}^{+\infty} \exp(i2\pi\nu x) * Ii(x) dx \\ bx(\nu) &= \int_{-\infty}^{+\infty} \exp(i2\pi\nu x) * BX(x) dx \end{aligned} \quad (12)$$

then we get the following relation for the Fourier transform $SL(\nu)$ of the truncated interferogram $IL(x)$

$$\begin{aligned} SL(\nu) &= \int_{-\infty}^{+\infty} \exp(i2\pi\nu x) * IL(x) dx \\ &= \int_{-\infty}^{+\infty} \exp(i2\pi\nu x) * Ii(x) BX(x) dx \\ &= \int_{-\infty}^{+\infty} Si(k) bx(\nu - k) dk \\ &= Si(\nu) O bx(\nu) \end{aligned} \quad (13)$$

The computation of a convolution integral $B(\nu) O C(\nu)$ as in Equation (13) can be visualized by the following procedure:

- Put the function B onto the k -axis with its origin at $k = 0$. This is $B(k)$. Do the same for function $C(k)$.
- Move the origin of function $C(k)$ to another position $k = \nu$ and reflect it

about this position. This yields $C(\nu - k)$.

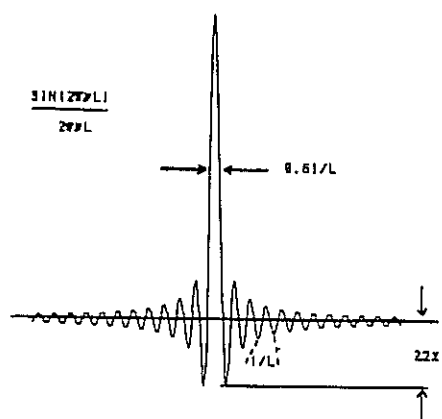
- Multiply the displaced and reversed function $C(\nu - k)$ by $B(k)$ and measure the area under the product function. This is the value of the convolution integral for one position.
- Repetition of the above three steps for all positions yields the complete wavenumber dependence of the convolution integral.

According to Equation (13), the spectrum $SL(\nu)$ of a finite interferogram can thus be obtained by convolving the spectrum $Si(\nu)$ corresponding to infinite optical path difference (and hence to infinite resolution, see part 1) with the 'instrumental lineshape' (ILS) function, $bx(\nu)$. This enables a clear description of a measured lineshape in terms of a natural lineshape (NLS) due to physical line broadening, the ILS representing solely the contribution of finite resolution.

The analytical form of the ILS corresponding to boxcar truncation can be easily derived from the Fourier integral of a unity operand using a finite integration range. The result is the well-known sinc function:

$$\begin{aligned} bx(\nu) &= L * \text{sinc}(2\nu L) \\ &= L * \sin(2\pi\nu L) / (2\pi\nu L) \end{aligned} \quad (14)$$

which is plotted in Figure 2. One sees, besides a main maximum centered about $\nu = 0$, numerous additional peaks, called side lobes or 'feet'. These side lobes



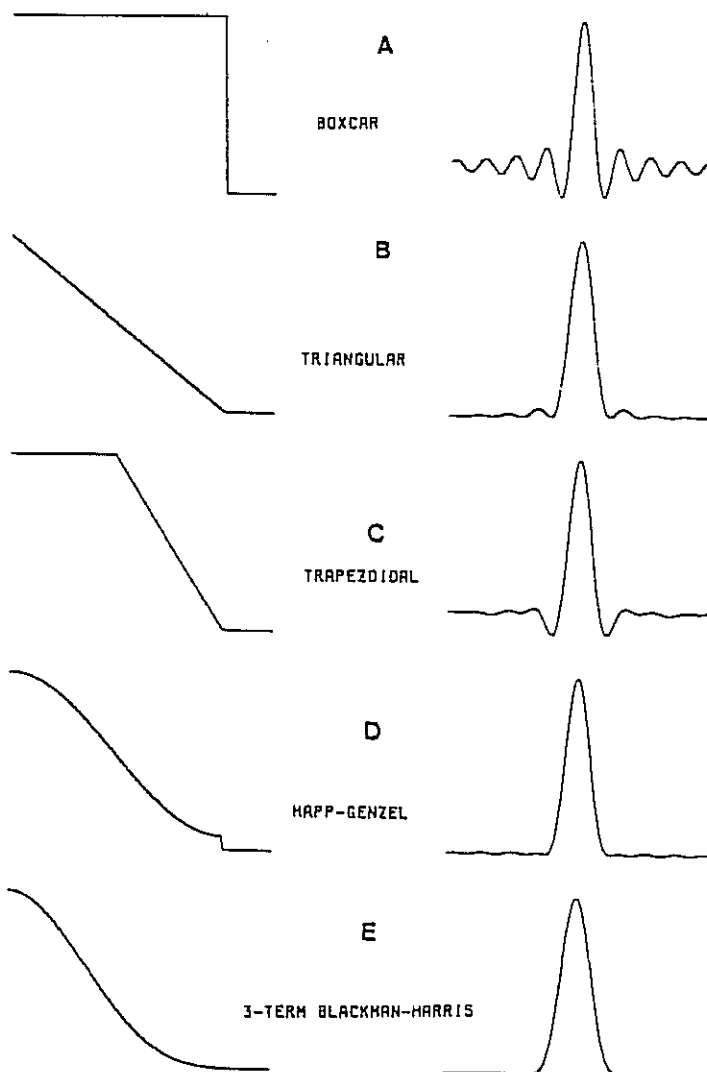
cause a 'leakage' of the spectral intensity, i.e. the intensity is not strictly localized but contributes also to these side lobes. The largest side lobe amplitude is 22% of the main lobe amplitude.

As the side lobes do not correspond to actually measured information but rather represent an artifact due to the abrupt truncation at $x=L$, it is desirable to reduce their amplitude. The process, which attenuates the spurious 'feet' in the spectral domain, is known as 'apodization' (originating from the Greek word *αποδο*, which means 'removal of the feet').

2.2 Solution to Leakage: Apodization

The solution to the problem of leakage i is to truncate the interferogram less abruptly than with the rectangular or 'boxcar' cutoff. This is equivalent to finding a cutoff or apodization function with a Fourier transform which shows fewer side lobes than the sinc function. Numerous such functions exist. An extensive overview of their individual properties can be found in the review by Harris [2]. In Figures 3B - 3D, four examples of such functions and their Fourier transforms are plotted together with the boxcar cutoff in Figure 3A. The analytical forms of these apodization functions are

$$\begin{aligned} \text{Trinagular (TR):} \\ \text{TR}(x) &= 1 - n/L \quad (n = 0, 1, \dots, L) \end{aligned} \quad (15)$$



Trapezoidal or Four - point (FP) :

$$FP(x) = 1 \quad (n = 0, 1, \dots, BPC) \\ 1 - [n - BPC] / [BPC - BPC] \quad (16) \\ (n = BPC, \dots, BPD)$$

This is a boxcar function between 0 and breakpoint BPC , and a triangular function between breakpoints BPC and BPD . In Figure 3C we choose $BPD = L$.

Hamming or Happ - Genzel (HG) :

$$HG(x) = 0.54 + 0.46 \cos(\pi n / L) \quad (17) \\ (n = 0, 1, \dots, L)$$

This is a cosine halfwave on a boxcar pedestal. The amplitude at the boundary $x = L$ is not zero but still 8% of the amplitude at the origin. The parameters 0.54 and 0.46 have been chosen for optimum suppression of the first, largest side lobe.

Three and four - term Blackmann - Harris (BH) :

$$BH(x) = A0 + A1 \cos(\pi n / L) \\ + A2 \cos(\pi 2n / L) \\ + A3 \cos(\pi 3n / L) \quad (18) \\ (n = 0, 1, \dots, L)$$

This set of windows is a generalization of the Happ-Genzel function. The coefficients have been optimized numerically to trade main lobe width for side lobe suppression (see [2]):

	3-term BH	4-term BH
A0	0.42323	0.35875
A1	0.49755	0.48829
A2	0.07922	0.14128
A3	0.0	0.01168

The three-term BH-window is plotted in Figure 3E.

2.3 Apodization and Resolution

As expected, Figure 3 reveals that all apodization functions produce an ILS with a lower side lobe level than the sinc function. However, one also sees that the main lobes of all ILS's in Figures 3B - 3E are broader than that of the sinc function in Figure 3A. The width at half height (WHH) of the ILS defines the best resolution achievable with a given apodization function. This is because if two spectral

lines are to appear resolved from one another, they must be separated by at least the distance of their WHH, otherwise no 'dip' will occur between them. As side lobe suppression always causes main lobe broadening, leakage reduction is only possible at the cost of resolution.

The choice of a particular apodization function depends therefore on what one is aiming at. If the optimum resolution of $0.61/L$ is mandatory, boxcar truncation (= no apodization) should be chosen. If a resolution loss of 50% compared to the boxcar can be tolerated, the Happ-Genzel-, or even better, the 3-term BH-apodization is recommended. If the interferogram contains strong low-frequency components, it may show an offset at the end, which would produce 'wiggles' in the spectrum. To suppress these wiggles, one should use a function that is close to zero at the boundary, such as the triangular-, trapezoidal-, or Blackman-Harris windows. As the ILS produced by the Blackman-Harris function shows nearly the same WHH as the triangular- and Happ-Genzel function (roughly $0.9/L$), but at the same time, the highest side lobe suppression and is furthermore nearly zero at the interval ends, it can be considered the top performer of these three functions.

In practice, the shape of a spectral line measured at finite resolution is always a mixture of natural and instrumental lineshape. This is demonstrated in Figures 4A - 4C, where the same NLS (in our case a Lorentzian), recorded at different resolutions, is plotted. The ILS corresponds to boxcar truncation. One sees that a lineshape close to the NLS can only be observed if the width of the ILS is small compared to the NLS.

At the end of the discussion of apodization, it should be noted that an instrumental lineshape with side lobes is of course also imposed on spectra from dispersive instruments. The ILS produced by the slit of a grating spectrometer corresponds to the ILS caused by triangular apodization. The difference between FT-IR and dispersive spectroscopy concerning apodization is that an FT-IR spectroscopist can choose the optimum ILS for his specific needs, while the 'dispersive spectroscopist' cannot.

3 Phase Correction

The last mathematical operation to be performed during the conversion of an interferogram into a spectrum is phase correction.

Phase correction is necessary, because the FT of a measured interferogram generally yields a complex spectrum

$C(\nu)$ rather than a real spectrum $S(\nu)$ as known from conventional spectrometers. A complex spectrum $C(\nu)$ can be represented by the sum

$$C(\nu) = R(\nu) + iI(\nu) \quad (19)$$

of a purely real part $R(\nu)$ and a purely imaginary part $I(\nu)$ or, equivalently, by the product

$$C(\nu) = S(\nu) \exp(i\phi(\nu)) \quad (20)$$

of the true 'amplitude' spectrum $S(\nu)$ and the complex exponential $\exp(i\phi(\nu))$ containing the wavenumber-dependent 'phase' $\phi(\nu)$.

The aim of the phase correction procedure is to extract the amplitude spectrum $S(\nu)$ from the complex output $C(\nu)$ of the FT. This can be done either by calculating the square root of the 'power spectrum' $P(\nu) = C(\nu) * C^*(\nu)$:

$$S(\nu) = [C(\nu) * C^*(\nu)]^{1/2} \\ = [R^2(\nu) + I^2(\nu)]^{1/2} \quad (21)$$

or by multiplication of $C(\nu)$ by the inverse of the phase exponential and taking the real part of the result:

$$S(\nu) = \text{Re}[C(\nu) \exp(-i\phi(\nu))] \quad (22)$$

The phase $\phi(\nu)$ in the exponential $\exp(i\phi(\nu))$ can be computed from the relation

$$\phi(\nu) = \arctan[I(\nu) / R(\nu)] \quad (23)$$

Equations (21) and (22) are equivalent, if one deals with perfect data free of noise. However, if noise is present, as is always the case with measured data, noise contributions computed from Equation (21) are always positive and in the worst case, a factor or $\sqrt{2}$ larger than the correctly signed noise amplitudes computed from Equation (22). Procedure (22) is known as 'multiplicative phase correction' or the 'Mertz method' [3].

3.1 Reasons for the Non-Zero-Phase

The reason for getting a complex spectrum out of the FT is that the input to the FT is not mirror-symmetrical about the point $x = 0$. The asymmetry of the FT input originates from three different sources:

- None of the sampling positions coincides exactly with the proper position of zero path difference. This is generally the case and causes a phase linear in ν .
- Only a 'one-sided' interferogram is measured, i.e. only one side is recorded to its full extent, the other consists only of a few hundred points.
- The interferogram may be 'intrinsically' asymmetric. This may be due to wavenumber-dependent phase delays of, either the optics, the detector/amplifier unit, or the electronic filters.

Figure 5 shows how the phase contributions from a) and b) can be calculated from a short double-sided portion of length PIP ($PIP/2$ points on either side of the centerburst). This yields, after apodization, zero filling, and FT gives a low resolution phase spectrum. As the phase is a slowly varying function of ν (except in the regions of beamsplitter or filter cutoff), a low resolution phase spectrum is sufficient. During the phase correction this is expanded to the size of the full array by interpolation.

The steps in the computation of the full array are shown in Figure 6. After apodization, the short double sided part is multiplied by a ramp to avoid counting it twice. The effect of this ramp can be understood by decomposing it into a boxcar of height 0.5 and a ramp through the origin, as indicated in Figure 6B. Because of the symmetry properties of the FT, the even boxcar function contributes only to the real part of the FT and has the effect of multiplying all contributions from $-PIP/2$ to $+PIP/2$ by $1/2$, while the odd part of the ramp contributes only to the imaginary part of the FT. The odd part of the ramp is added to avoid a step at $+PIP/2$, which would produce 'wiggles' in the spectrum.

Having multiplied the data array by the ramp and the apodization function, it is zero filled and Fourier transformed. Using Equation (22) the resulting complex array $C(\nu)$ is then multiplied by the complex exponential $\exp(i\phi(\nu))$, which is computed by interpolation from the low resolution phase spectrum.

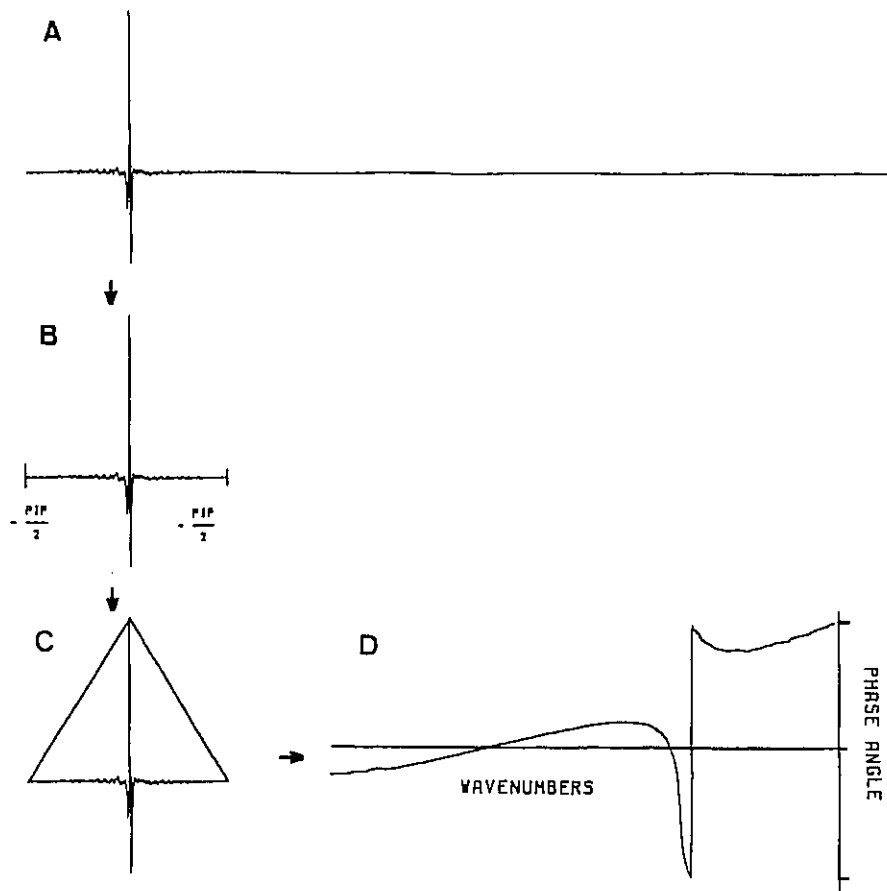
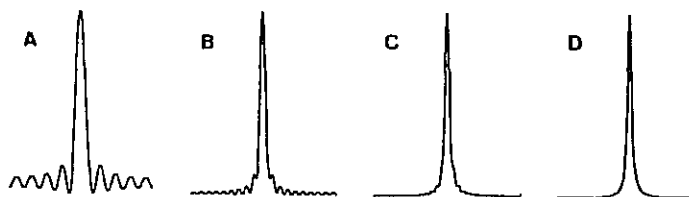
After the phase correction, the corrected real part $R'(\nu)$

$$R'(\nu) = R(\nu)\cos(\phi(\nu)) - I(\nu)\sin(\phi(\nu))$$

represents the final single-channel

spectrum $S(\nu)$ which is stored for further processing. The corrected imaginary part $I'(\nu)$

$$I'(\nu) = R(\nu)\sin(\phi(\nu)) + I(\nu)\cos(\phi(\nu))$$



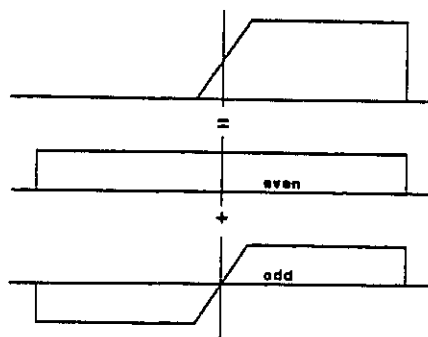
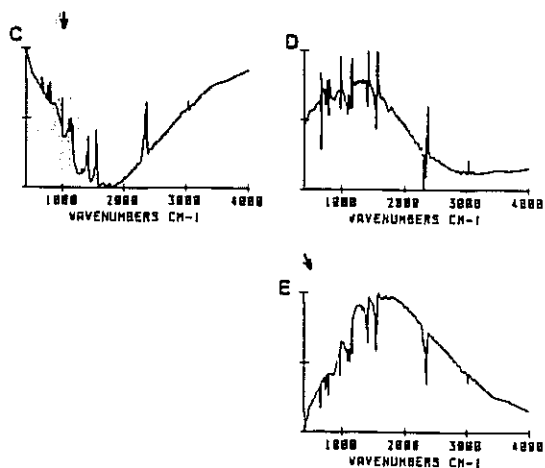
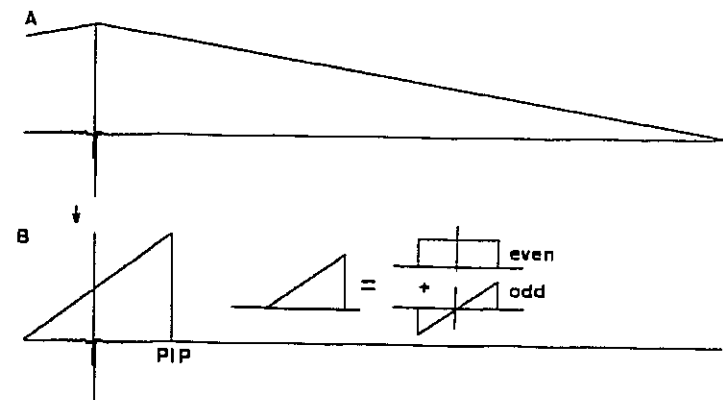
originates from the antisymmetric contribution of the cutoff at $-PI/2$ and the ramp as is shown in Figure 7. $I'(\nu)$ would be zero, if a double-sided interferogram had been used. The corrected imaginary part is normally skipped or not even calculated to save computer time. The fact that $I'(\nu)$ is non zero after the phase correction demonstrates that the

direct calculation of $S(\nu)$ from the power spectrum via Equation (21) will yield erroneous results in the case of one-sided interferograms, because this procedure corrects for all contributions to $I(\nu)$, including those from the cutoff at $-PI/2$. This method is only properly applicable to double-sided interferograms. With this last topic we have now completed the discussion of the standard operations necessary to convert an

interferogram into its spectrum. In the next installment, we will deal with several FT- or interferogram-based techniques, which are very useful in IR- and GC-IR spectroscopy.

References

- [1] W. Herres and J. Gronholz, *Comp. Appl. Lab.* 2 (1984) 216.
- [2] F.J. Harris, *Proceedings of the IEEE*, 66 (1978) 51.
- [3] L. Mertz, *Infrared Phys.* 7 (1967) 17.



UNDERSTANDING FT-IR DATA PROCESSING PART 3: FURTHER USEFUL COMPUTATIONAL METHODS

In Parts 1 and 2 of this series we covered the standard operations of FT-IR, from data acquisition to the final spectrum. This third and last part continues with a discussion of additional useful techniques and closes with a remark about the data system.

In Parts 1 and 2 [1, 2] of this series we dealt with data acquisition in a Fourier transform infrared (FT-IR) spectrometer and described the processing of the raw data to generate the final spectrum. To perform the necessary calculations in a reasonable time, the minicomputer of an FTIR spectrometer needs sufficient on-line computing power and is thus also very well suited to do more than just calculation of FFT's, phase correction, and ratioing of spectra. In fact, the considerable inherent number-crunching capability inside an FT-IR spectrometer is sometimes called the fourth advantage of FTIR. In this last part of our series we discuss some examples of such additional data processing in both domains: the frequency domain and the interferogram domain.

1 The Problem of "Ghost" Interferograms or Fringes

The appearance of sinusoidal modulations, called 'fringes' on the baseline of IR-spectra is well known to spectroscopists. These fringes or 'channel spectra' result from multireflections of the IR beam between the surfaces of a plane-parallel device in the spectrometer's

optical path, such as the plane parallel sample itself or a liquid-cell window. They can disturb the 'useful' spectral information quite seriously, as is shown for the absorbance spectrum of a silicon wafer in Figure 1A. From Figure 2 in [1] it is readily concluded that fringes of constant frequency in the wave number domain must result from a single 'spike' or a narrow 'signature' in the corresponding interferogram. Indeed, the sample interferogram from which the absorbance spectrum in Figure 1A was calculated, shows a 'ghost' interferogram or 'echo peak' at an offset of 4206 interferogram points from the centerburst (see Figure 1B). The interferogram was acquired at a resolution of 2 cm^{-1} and thus consists of 8192 points. Fourier transformation of the full interferogram including the echo peak yields the spectrum in Figure 1A, whereas reduction of the resolution to 4 cm^{-1} (equivalent to an interferogram of 4096 data points) truncates the interferogram just before the echo peak yielding the 'fringe-free' spectrum of Figure 1C.

This example shows that one can easily get rid of the fringes by discarding all interferogram points from the echo peak to the end. This can be achieved either by choosing a lower optical resolution during data acquisition or by apodization with a trapezoidal function (see Figure 3 of [2]) chosen such that all points from the echo peak to the end are set to zero.

1.1 Mathematical Description of the Fringes

The single channel spectrum $S(\nu)$ of the IR-radiation transmitted through a multireflecting plate can be calculated from the empty-channel background spectrum $B(\nu)$ the half-space power reflectance $R(\nu)$ the refractive index $n(\nu)$, the absorption coefficient $a(\nu)$, the phase change $\vartheta(\nu)$ occurring at every internal reflection and the thickness d of the plate. The result is the well-known Airy formula [3] which we have extended to include absorptive intensity loss and cast into a form better suited for Fourier transformation:

$$S(\nu) = S + 2S_0 \sum_{k=1}^{\infty} \{R(\nu) \exp(-ad)\}^k \times \cos(4\pi kn(\nu)d + 2k\vartheta(\nu)) \quad (1)$$

with

$$S_0(\nu) = B(\nu) \frac{(1 - R(\nu))^2 \exp(-ad)}{1 - R^2(\nu) \exp(-2ad)} \quad (2)$$

Equation (1) shows nicely that the resulting spectrum can be represented by the sum of a fringe-free spectrum $S_0(\nu)$ and an infinite number of interference terms. Due to the factor $(R(\nu) \exp(-ad))^k$,

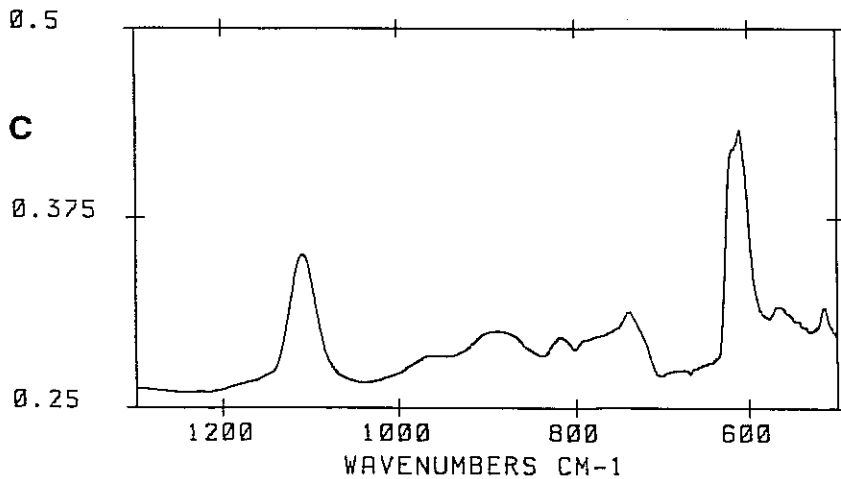
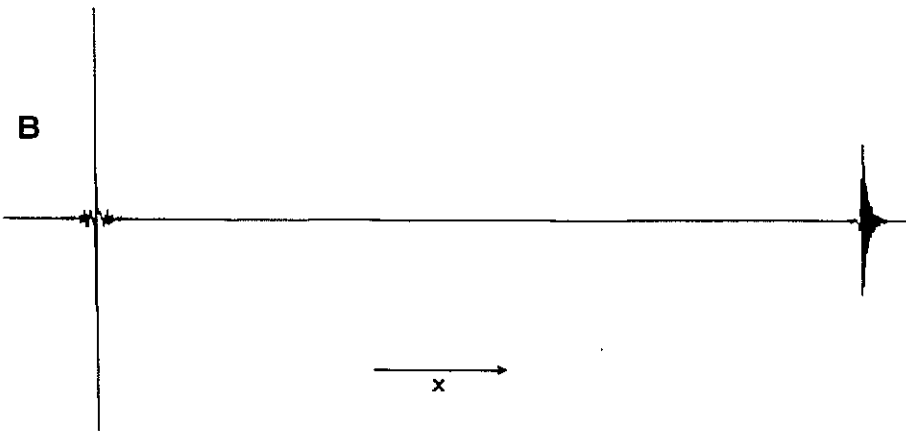
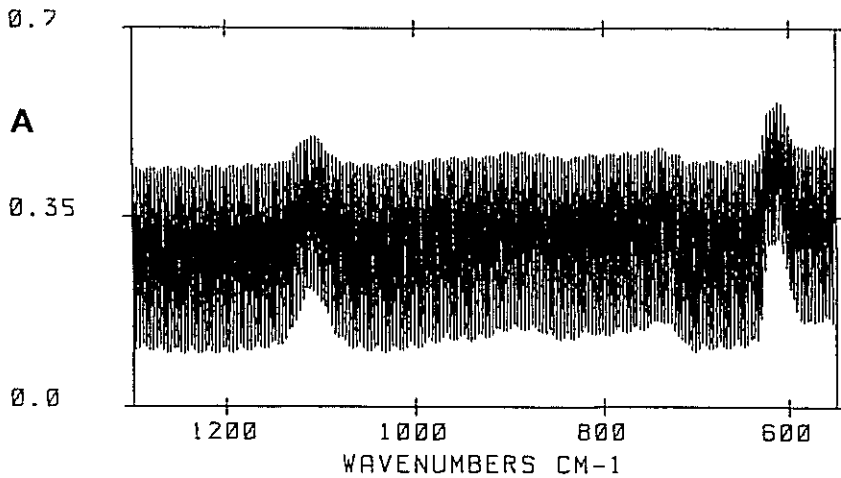


Figure 1: A) Absorbance spectrum of a silicon wafer calculated using the full length (8K points = RES 2cm-1) of the interferogram from Fig. 1B. B) Sample interferogram of a silicon wafer. Spectral resolution corresponds to 2 cm⁻¹ (= 8K points). The distance between the small 'ghost' interferogram on the right and the centerburst is 4206 pints (y-scale expanded for clarity). C) Absorbance spectrum of a silicon wafer calculated from the interferogram of Fig. 1B using only 4K data points (resolution = 4 cm⁻¹).

the intensity of the k -th order interference term decreases with ascending order k because $R^k(\nu)$ is always less than 1.

The interferogram corresponding to $S(\nu)$ can be calculated from eq. (1) by inverse Fourier transformation. For the sake of simplicity we assume that the refractive index $n(\nu)$ and the phase $\vartheta(\nu)$ within the argument of the cosine interference terms do not depend on ν . For this simplified case one gets

$$I(x) = I_0(x) + 2 \sum_{k=1}^{\infty} I_k(x) \circ [x - (2nk d + 2k\vartheta)] + 2 \sum_{k=1}^{\infty} I_k(x) \circ I_0[x + (2nk d + 2k\vartheta)] \quad (3)$$

This result shows that the cosine interference terms of the spectral domain correspond to additional ghost interferograms in the interferogram domain, appearing symmetrically on both sides of the centerburst at distances $2nk d + 2k\vartheta$. The shape of these echo peaks is given by convolution of the undisturbed main interferogram $I_0(x)$ with the 'interferogram' $I_k(x)$ which is the product of the reflectance spectrum $R(\nu)$ and the transmission spectrum $T(\nu) = \exp(-ad)$. For the same reason as given above, the intensity of the k th ghost interferogram is at least smaller by a factor R^k than the main centerburst. Due to the wavenumber dependence of $n(\nu)$ (i.e.: to dispersion) and of $\vartheta(\nu)$ which we explicitly neglected in the derivation of eq. (3), the echo peaks are generally additionally broadened and distorted and may be thus further reduced.

1.2 Use of Fringes for Thickness Determination

Eq. (3) shows, that the distance between the main centerburst and the echo peaks is directly proportional to the product of refractive index n and thickness d . If the average refractive index n is known, one can therefore calculate the thickness of the fringe-producing element. This possibility is, e.g., often used in semiconductor quality control to determine the thickness of epitaxial layers deposited on doped substrates.

Conversely, if both n and d are known, the offset X between main peak and first echo peak can be calculated as:

$$X = N * \Delta x$$

$$= 2 * n * d$$

or in points

$$N = 2 * n * d (\Delta x)^{-1} \quad (4)$$

with $(\Delta x)^{-1} = 157800 \text{ cm}^{-1}$ for mid-IR bandwidth, 0 – 7900 cm^{-1} .

The problem of ghost interferograms is of course not restricted to silicon wafers but can also be encountered in measuring standard alkali halide pellets or cast films on, e.g., KBr crystals. Sometimes spectra look 'noisy' at longer wavelength whereas expansion reveals fringes resulting from 'too perfect' a preparation, i.e. from plane-parallelism of the pellet.

In Table 1 the distance N between the interferogram centerburst and the first echo peak has been compiled for several thicknesses of four substrates commonly used in FT-IR. This table also shows the maximum resolution, which does not include the echo peak. It may be helpful in finding the proper combination of sample thickness and resolution for a given substrate.

1.3 Possibilities for Eliminating Fringes

The question might arise whether there is some method of eliminating all kinds of fringes, which plague the spectroscopist when measuring polymer films. Figure 2A shows the transmission spectrum of some technical polyethylene (PE)-type film. Although the modulations are very pronounced in the spectrum, the corresponding echo peak in the sample interferogram of Figure 2B can hardly be detected, being due to a much smaller reflectance $R(\nu)$ of PE as compared to silicon and to stronger dispersion, which additionally broadens the echo peak and reduces its size. Furthermore, the echo peak is located only 267 points away from the centerburst because the film is only about fifty microns thick. For this latter reason, truncation of the interferogram (= reduction of resolution) is here obviously no solution because the resulting resolution would be too poor. In those cases where reduction of resolution cannot be applied, there are at least two

Table 1: Location of the echo peak as a function of substrate thickness.

Offset of the echo peak relative to the centerburst as a function of the thickness of four commonly used substrates. Last column contains the maximum possible resolution for which still no interferences occur assuming 250 points before and 2 * * N- 250points after the centerburst in all cases. Bandwidth 7900 cm^{-1} -throughout. Value of refractive index n corresponds to 1000 cm^{-1} -

Substrate	Thickness [cm]	Offset from Centerburst [Points]	Maximum resolution without interferences [cm^{-1}]
KBr ($n = 1.53$)	0.05	2417	8
	0.1	4834	4
	0.15	7252	4
	0.2	9669	2
	0.3	14504	2
AgCl ($n = 1.98$)	0.05	3128	8
	0.1	6257	4
	0.15	9385	2
	0.2	12514	2
	0.3	18770	1
ZnSe ($n = 2.40$)	0.05	31284	1
	0.05	3792	8
	0.1	7584	4
	0.15	11386	2
	0.2	15168	2
Si ($n = 3.42$)	0.3	22752	1
	0.5	37920	0.5
	0.05	5403	4
	0.1	10807	2
	0.15	16211	1
	0.2	21614	1

other solutions to the fringe problem: an experimental and a numerical solution.

1.3.1 Experimental Solution of the Fringe Problem

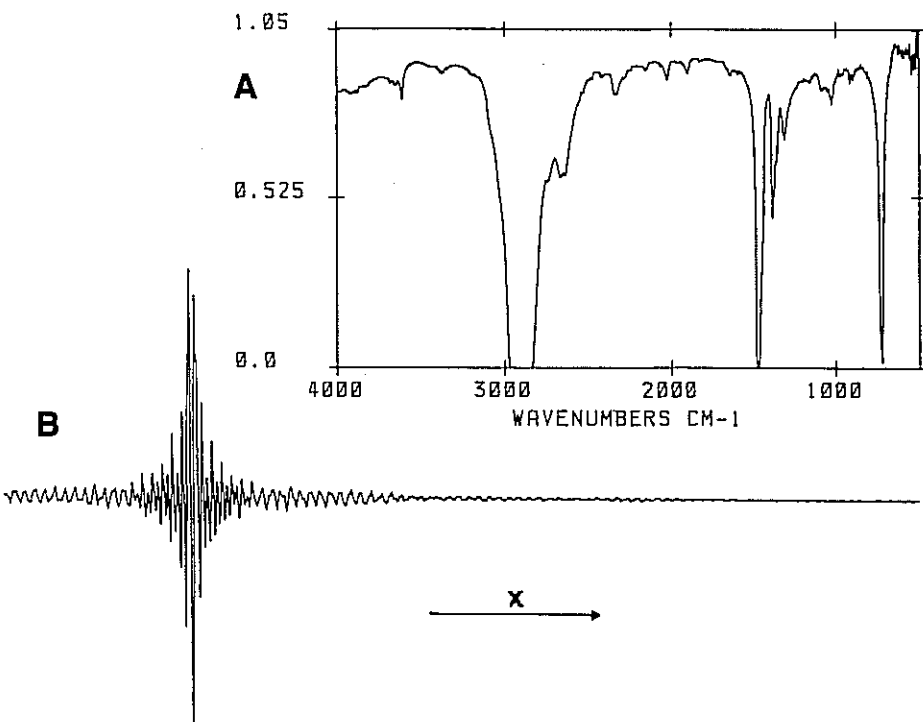
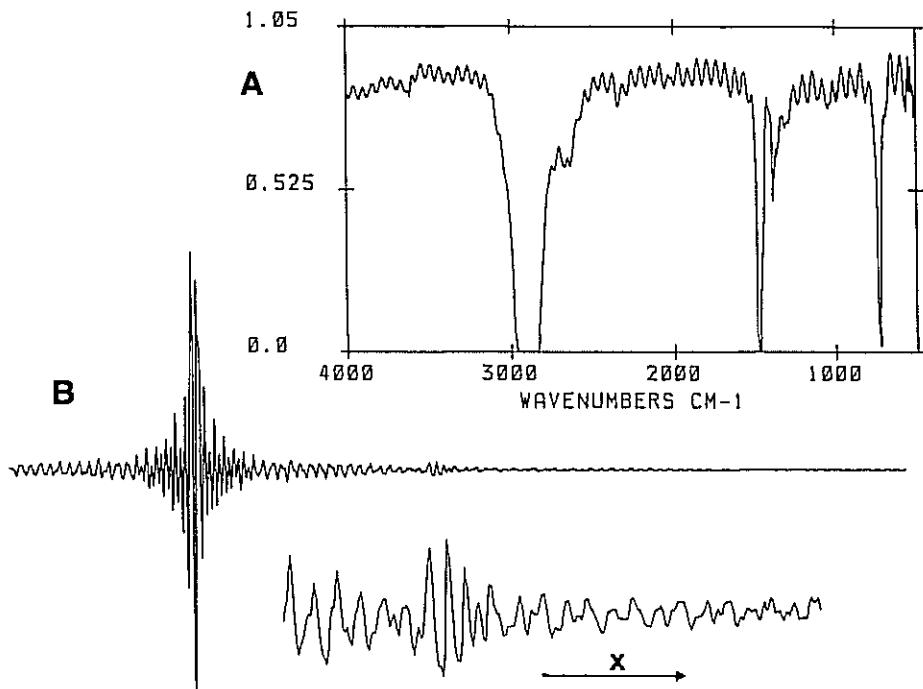
From Eq. (1) it is clear that all interference terms vanish if the reflectance $R(\nu)$ of the sample can be made zero. Experimentally, this is indeed possible using polarized IR radiation, as is shown in Figure 3A, where the same polymer film was measured. A KRS-5 polarizer was used to provide an IR beam polarized parallel to the plane of incidence, while the film itself was oriented to the incident light at the Brewster angle. As the reflectance of this type of polarized radiation approaches a minimum at the Brewster angle (or is ideally zero), multireflections and thus 'channel spectra' are also

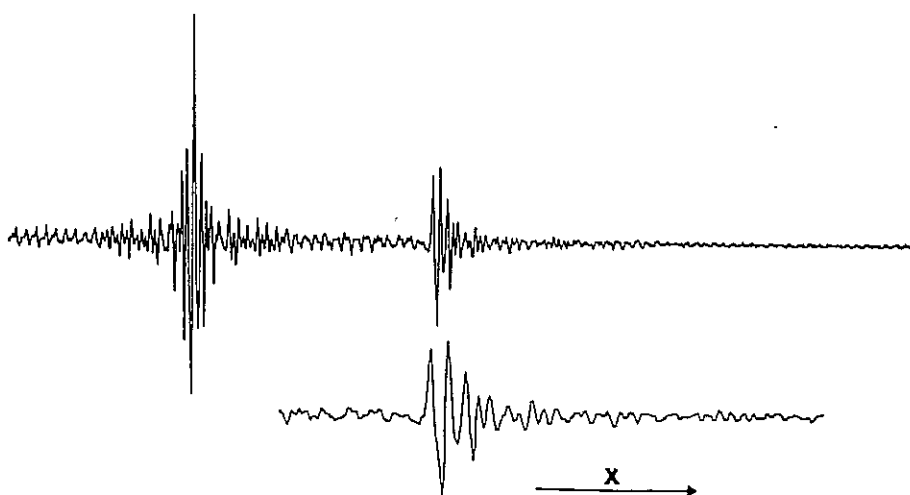
minimized. This experiment makes it possible to obtain a difference interferogram (Figure 2B minus Figure 3B) which reveals the structure of the echo peak more clearly (see Figure 4). Another experimental solution to the fringe problem would be a measurement using ATR (attenuated total reflection). However, unlike the Brewster angle measurement, ATR would change the relative intensities due to the wavelength-dependent depth of penetration.

1.3.2 Numerical Solution of the Fringe Problem

The undesired oscillations due to multireflections are spread over a large part of the spectrum but are confined to a small region in the interferogram domain. Numerical correction of the fringes is therefore easier in the interferogram domain. As we have learned from Eqs. (1) and (2), removal of the fringes from the spectrum is equivalent to removal of the echo peaks from the interferogram. It has been shown by Hirschfeld and Mantz [4] that this can be done by substituting the regions around the echo peaks, by zeroes, by a straight line, or by another reasonable guess of the undisturbed 0-th order interferogram. Using the interferogram of Figure 2B as an example, the substitution of the echo peak by a straight line is demonstrated in Figures 5A-C. One sees that, in contrast to the 'clean' experimental solution of Figure 3A, not all oscillations are removed but they are drastically reduced in size such that small peaks which were hidden under the oscillations are now clearly detectable. Although substitution by a straight line is a brute-force method, the result is quite useful but must always be regarded with caution as additional artifacts, like creation or cancellation of line splittings, can easily be introduced.

This numerical solution of the fringe problem can only be applied if the echo peak is intense enough to be detected and if it is far enough away from the centerburst.





While the smallness of the echo peak can be overcome by calculating its position from Eq. (5) if both n and d are known, violation of the second condition leads to severe baseline distortions and cannot be recommended.

Before leaving the interesting field of ghost interferograms, it is worthwhile noting that a 'simple' absorbance measurement of, say, a KBr pellet is not adequately described by the familiar formula

$$T(\nu) = \exp(-a(\nu)d)$$

with transmittance $T(\nu)$ and absorbance $a(\nu)d$, but must even in the simplest case of no multireflections be substituted by the numerator of Eq. (2).

$$T(\nu) = (1 - R(\nu))^2 \exp(-a(\nu)d) \quad (5)$$

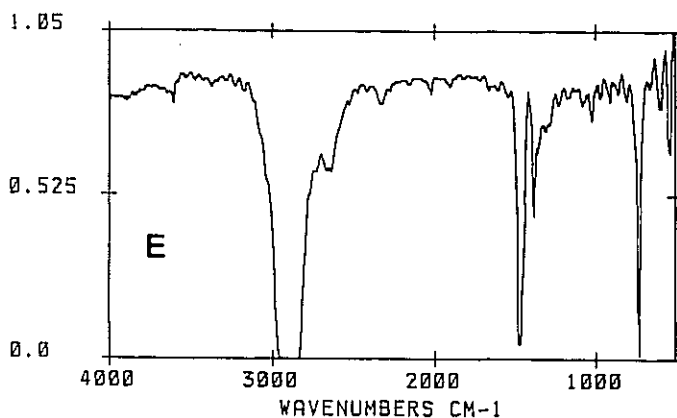
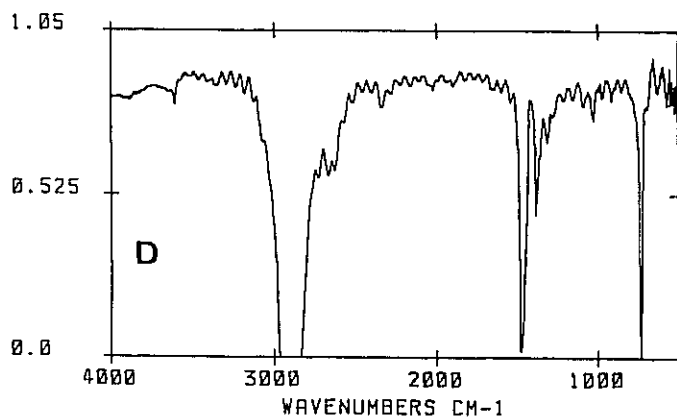
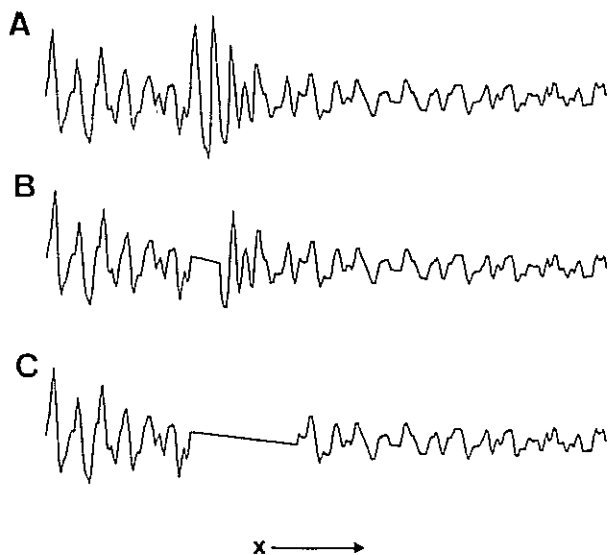


Figure 5: A) Part of the sample interferogram from Fig. 2B containing the signature. B) Same part of interferogram as in Fig. 5A, but points 466 to 479 (14 points) are substituted by a straight line to suppress the signature. C) Same as Fig. 6B but points 466 to 514 are substituted (49 points). D) Transmission spectrum calculated from the interferogram of Fig. 5B. E) Transmission spectrum calculated from the interferogram of Fig. 5C.

This means that one must not forget at least two reflections at both boundaries of the sample. The factor $(1 - R)^* * 2$ causes an intensity reduction of the transmitted radiation, but may also lead to changes in position and intensity of single lines if $R(\nu)$ is significantly structured. If the pellet has plane-parallel surfaces, multireflection comes into play and the full Eq. (2) (divided by the empty channel background spectrum $B(P)$) must be used. Thus, from a transmission measurement one always gets a mixture of reflection and absorption properties unless the reflectance contributions are explicitly corrected for. This single or multireflection correction, i.e. extraction of the absorbance $a(P)d$ from equations (5) or (2) is another task which could be routinely performed by the FT-IR spectrometer's minicomputer. It is mandatory if one is interested in precise determination of optical constants.

2 Interrelation of Smoothing and Apodization

In Part 2 of this series we already described apodization at some length. Apodization means multiplication of the interferogram by a decaying function. Its effect on the spectrum is a suppression of side lobes at the expense of decreased resolution. We also showed that multiplication of the interferogram $I(x)$ by a function $a(x)$ is equivalent to convolution of the spectrum $S(\nu)$ by the Fourier transform $A(\nu)$ of the apodization function.

Three of the special apodization functions discussed (Happ-Genzel 3- and 4-term Blackman-Harris) consisted of a sum of cosine functions as

$$BH(x) = A0 + \sum_{n=1}^N An \cos(n * \pi * x / L).$$

It is instructive to calculate the Fourier transform of such a function and look at the corresponding convolution in the spectral domain in more detail. One gets

$$\begin{aligned} S'(n * \Delta \nu) &= A0 S(n * \Delta \nu) \\ &+ A1(S((n-1) * \Delta \nu) + S((n+1) * \Delta \nu)) \\ &+ A2(S((n-2) * \Delta \nu) + S((n+2) * \Delta \nu)) \quad (6) \\ &K \\ &+ AN(S((n-N) * \Delta \nu) + S((n+N) * \Delta \nu)) \end{aligned}$$

Hence, in order to calculate one point $S'(n * \Delta \nu)$ of the spectrum corresponding to the apodized interferogram one must:

- multiply the ordinate $S(n * \Delta \nu)$ of the non-apodized spectrum by $A0$,
- multiply the left and right next neighbors of $S(n * \Delta \nu)$ by $A1$,
- multiply the left and right second next neighbors by $A2$ and so forth (up to $N = 3$ for 4-term BR)
- sum the intermediate results. This is $S'(n * \Delta \nu)$.

This shows that the points of the spectrum $S'(n * \Delta \nu)$ are a weighted mean of adjacent points of the non-apodized case $S(\nu)$, the coefficients An of the apodization function being the weighting factors. As taking a weighted mean of spectral data amounts to nothing but smoothing, we conclude that apodization in the interferogram domain is equivalent to smoothing in the spectral domain. (Note that this special family of apodization functions would easily allow one to perform the apodization after the FT by summing neighboring points of the non-apodized spectrum multiplied by appropriate factors.)

Smoothing is an essential tool for reducing the noise in a spectrum. In addition to apodization, smoothing is mostly done by the famous Savitzky-Golay procedure [5], the merit of which is that (although a least-squares method and therefore data-dependent) the weighting coefficients used in the averaging process are fixed integers. This means that smoothing can be programmed in integer arithmetic leading to short computation times even on machines without floating point hardware support. As spectral smoothing is so widespread and well known, we omit a detailed description and rather turn to the question of what happens if smoothing is done in the interferogram domain.

2.1 Smoothing in the Interferogram Domain: Digital Filtering

The effect of smoothing an interferogram on the corresponding spectrum is demonstrated in Figure 6A-C. Figure 6A shows an unmodified single channel spectrum, Figure 6B shows the spectrum corresponding to the same interferogram smoothed using a 9-point Savitzky-Golay window. Figure 6C is the ratio of Figures 6B and 6A. One sees that the effect of smoothing the interferogram is to decrease the intensity of the spectrum towards higher wavenumbers, i.e. it behaves like a low-pass filter with the frequency response of Figure 6C. Accordingly, convolution in the interferogram domain is termed 'digital filtering'.

Digital filtering is closely related to apodization although the aims of both procedures are completely different. While during apodization one multiplies the interferogram by the apodization function to achieve a certain line shape (convolution), digital filtering means that one convolves the interferogram by a filter function to achieve a certain (multiplicative) frequency response. In practice, the filter function is seldom of the Savitzky-Golay type (this was simply used as an example) but is derived from the desired frequency response by inverse Fourier transformation of the desired spectral bandpass in case of non-recursive filters.

The advantage of digital - compared to analog - filtering lies in its great flexibility, because almost any conceivable frequency response can be modeled. The only problem is that, even today, where high-speed multipliers/adders are available, the time needed for computing the necessary convolutions is often still too long for real-time filtering in high-speed applications with data rates > 100 KHz. For slow-scanning FT-IR spectrometers, however, digital filtering can be appropriate and is then another task to be performed by the spectrometer's minicomputer.

3 Deconvolution

Deconvolution is another interesting FT-based method closely related to apodization and convolution. The aim of deconvolution is to decrease the widths of all lines in a limited spectral region or, equivalently, to enhance the apparent resolution of the spectrum.

The method is based on the assumption that the investigated spectrum $S(\nu)$ may be represented by a convolution

$$S(\nu) = S'(\nu) \circ L(\nu) \quad (7)$$

of a deconvolved spectrum $S'(\nu)$ containing sharp lines and a function $L(\nu)$ which is responsible for the line broadening. From the discussion of apodization in Part 2 of this series we know that convolution in the spectral domain corresponds to simple multiplication in the interferogram domain. Thus, inverse FT of Eq. (7) yields the product

$$I(x) = I'(x) * \lambda(x) \quad (8)$$

from which the broadening effect of $L(\nu)$ can be easily cancelled by division by $\lambda(x)$. The deconvolved spectrum $S'(\nu)$ is then obtained by another forward FT to the spectral domain. $I(x)$, $I'(x)$, and $\lambda(x)$ represent the inverse Fourier transforms of $S(\nu)$, $S'(\nu)$, and $L(\nu)$, respectively. As an easy example we consider the case of an absorbance spectrum $S(\nu)$, consisting of a single Lorentzian line which may be represented by convolution of an infinitely sharp line (a delta function) at $\nu = \nu_0$

$$S'(\nu) = \delta(\nu - \nu_0) \quad (9)$$

with a line shape function

Figure 6: A) Single-channel sample spectrum of the polymer film used for Figs. 2, 3. B) Same spectrum as in Fig. 6A but after smoothing the interferogram by a 9-point Savitzky-Golay window. The intensity towards higher wavenumbers is decreased. C) Ratio of Fig. 6B and Fig. 6A, showing the frequency response of the Savitzky-Golay 9-point digital filter (see text).

$$L(\nu) = \frac{a/\pi}{a^2 + \nu^2} \quad (10)$$

$$S(\nu) = S'(\nu) \circ L(\nu) \\ = \delta(\nu - \nu_0) \circ \frac{a/\pi}{a^2 + \nu^2} \quad (11)$$

$$= \frac{a/\pi}{a^2 + (\nu - \nu_0)^2}$$

As was shown in Figures 2A, C of [1], a sharp line corresponds to a cosine function in the interferogram domain

$$I'(x) = \cos(2\pi\nu_0 x) \quad (12)$$

whereas a Lorentzian at wavenumber ν_0 corresponds to a cosine damped by (i.e. multiplied by) an exponential decay function

$$\lambda(x) = \exp(-2\pi a|x|) \quad (13)$$

and is thus represented by the product

$$I(x) = I'(x) * \lambda(x) \\ = \cos(2\pi\nu_0 x) \exp(-2\pi a|x|) \quad (14)$$

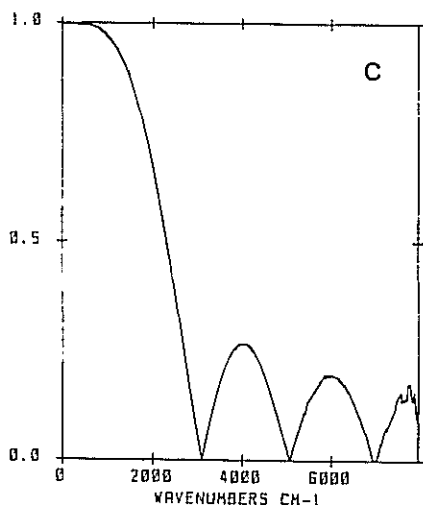
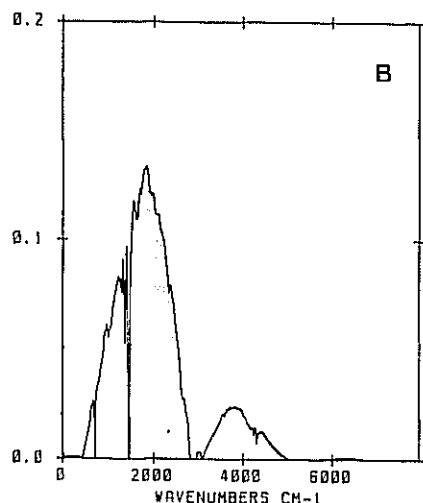
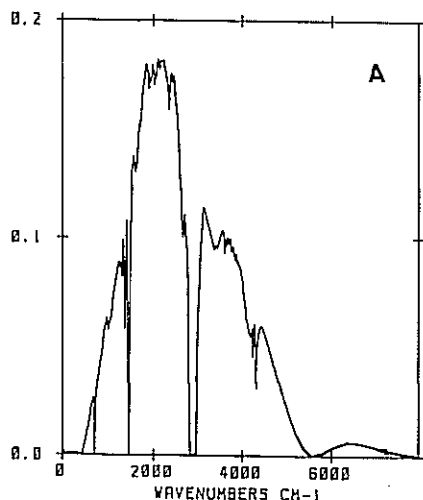
Deconvolution (= removal of the damping function $\lambda(x)$) is here obviously achieved by multiplication of $I(x)$ by the inverse of $\lambda(x)$

$$1/\lambda(x) = \exp(+2\pi a|x|) \quad (15)$$

yielding first the Fourier representation $I'(x)$ of the deconvolved spectrum $S'(\nu)$ and - after another forward FT - the deconvolved spectrum $S'(\nu)$ itself.

Applied to experimental data, deconvolution is less trivial than it appears from the synthetic case above because the proper form of the line shape function $L(\nu)$ is not known a priori. $L(\nu)$ can only be approximated by making a more or less reasonable assumption of its form (e.g. Lorentzian or Gaussian, or a mixture of both) and by estimating its width from the narrowest line in the investigated spectral area.

Another complication arises from the noise, which is always present in experimental data, as it is strongly amplified by the removal of the damping function. It has been shown by Kauppinen *et al.* [6] that noise effects can be partly suppressed by apodization of $I(x)$ by a



triangular function which truncates $I(x)$ at a point $x = E$ before the end at $x = L$. If it happens that the assumed $L(\nu)$ exactly equals the true $L(\nu)$, the shape of the lines in the deconvolved spectrum $S'(\nu)$ will be exactly of the type

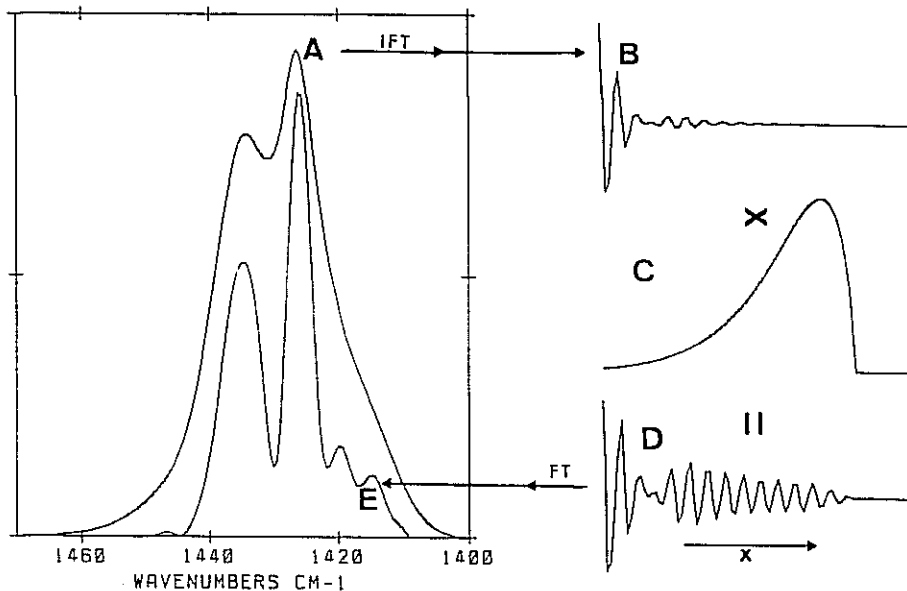
$$L'(\nu) = E \sin^2(\pi\nu E)$$

due to the apodization.

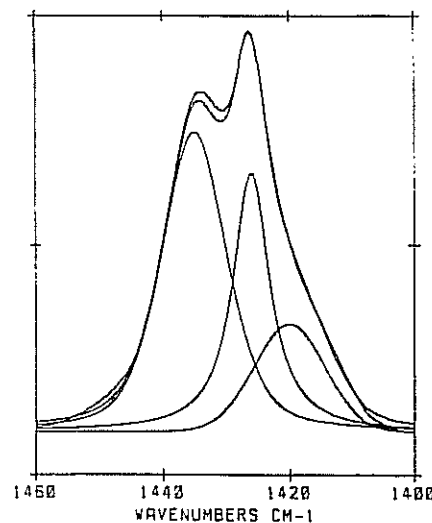
An example of what one typically may expect from deconvolution is given in Figures 7A-E. By an inverse FT, the original spectrum of Figure 7A is transformed to the interferogram domain (Figure 7B) and enhanced (multiplied) by the function of Figure 7C yielding Figure 7D which, after another forward FT, results in the resolution enhanced spectrum of Figure 7E. On comparing Figure 7A with Figure 7E, one sees that the original spectrum has been resolved into four components with an indication of a fifth component near 1450 cm^{-1} . The question, whether all of the smaller three components are true or artificial is not easy to solve. Additional oscillations may easily be introduced by 'over-deconvolution', i.e. by overestimating the widths of the line shape $L(\nu)$, as shown in Figure 9. In case of doubt, one should always underestimate the proper line width; this leads to lines broader than the achievable minimum but also with fewer artifacts.

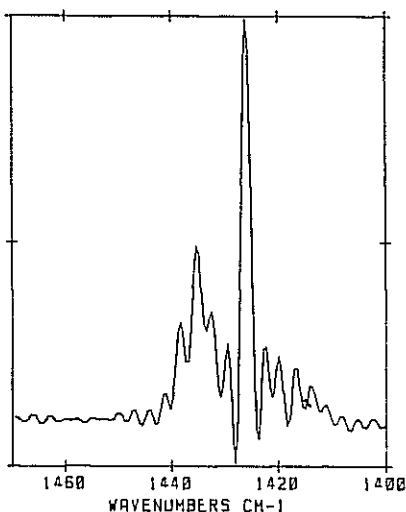
4 Spectrum Simulation

Deconvolution allows one to determine peak positions, relative intensities, and the number of individual components contributing to a certain spectral area. It should not be confused with spectrum simulation, which works in the spectral domain and which tries to represent a given spectrum by superposition of individual lines, the parameters of which (position, height, width, type) are optimized such that the deviation between experimental and simulated curve approaches a minimum. It should be noted, however, that deconvolution is well able to provide good starting values for the parameters of a spectrum simulation. This has been documented in Figure 8, which represents the result of an automatic least-squares spectrum simulation using the example of Figure 7A and the results from the deconvolution as starting values for the variation parameters. This shows that both methods complement each other nicely.



Deconvolution and spectrum simulation are mostly applied to absorbance spectra but they could equally well be used for analyzing other data like chromatograms.





ability of an additional FT- or array processor.

5.2 GC Traces Directly from the Interferogram

Besides window GC traces also non-frequency selective 'total' IR chromatograms similar to a GCFID trace can be calculated in various ways [7,8]. From the algorithms working directly on the interferometric data, a vector projection technique using the Gram-Schmidt orthonormalization procedure became fairly popular as 'Gram-Schmidt' technique and shall be treated below. Extracting chromatographic information directly from the interferogram without FT is fast and thus also possible on minicomputers without dedicated FT-processor.

A segment of each reference interferogram of N points is extracted, treated as an N dimensional vector r_n , and used to construct a set of $M < N$ orthonormal basis vectors b_n . This set of basis vectors represents the starting condition of pure carrier gas.

During the subsequent GC run the same segment from each acquired interferogram is used as a sample vector s . This sample vector is projected onto the orthonormal set of basis vectors and compared to its projection p by taking the vector difference $s - p$. This procedure is explained in Figure 11 for the case of just two basis vectors. While p represents the part of s due to pure carrier gas, the difference $s - p$ represents deviations therefrom, i.e. the value of $s - p$ constitutes the actual point of the chromatogram trace.

5.2.1 Gram-Schmidt Technique

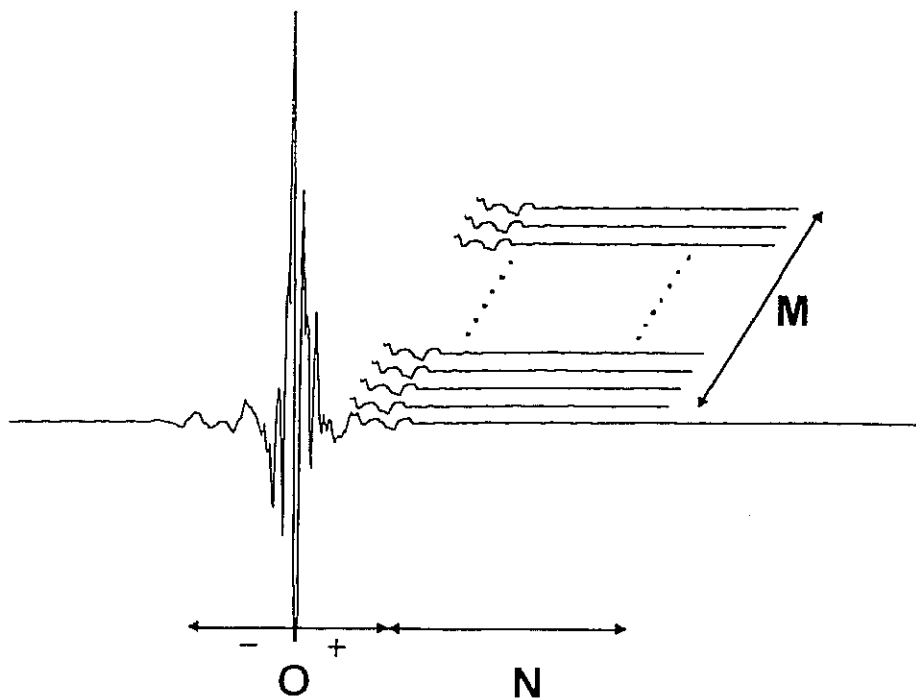
In the Gram-Schmidt technique a reference set of M interferograms is collected directly before the GC run when only carrier gas is leaving the GC column.

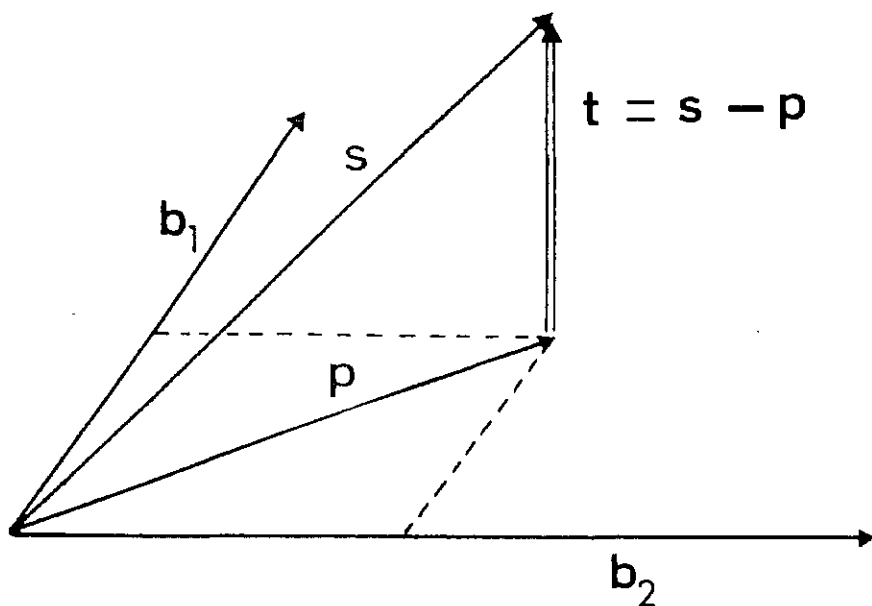
5 Generation of GC-IR Chromatograms

The combination of fast scanning spectrometers with gas chromatography (GC) and especially with capillary gas chromatography (HRGC) has evolved to the powerful hyphenated techniques GC-FTIR and HRGCFTIR. In these techniques, the FTIR spectrometer is used to measure complete interferograms of the gas leaving a GC-column at constant time intervals. From the acquired interferograms different kinds of chromatographic traces may be generated.

5.1 Spectral Window Chromatograms

Spectral window chromatograms monitor the change in absorbance in discrete spectral regions. This calculation involves all steps of computation typical for FT-IR spectroscopy like apodization, FT, phase correction, and ratioing to a stored background spectrum and is therefore a demanding task for the spectrometer's minicomputer. If it is to be done in real-time, with both high time resolution and sufficient spectral resolution, it generally needs the number crunching ca-





A practical example of the Gram-Schmidt trace is shown in Figure 12 in comparison with a conventional GC-FID trace. One sees the excellent agreement between the two chromatograms and also the excellent sensitivity of the Gram-Schmidt method. The achievable sensitivity depends on the setting of the Gram Schmidt parameters, namely the number of points N per vector, their offset O relative to the centerburst and the number $M < N$ of basis vectors (compare Figure 10). The optimization of these parameters is still a field of active research [9,10]. A good parameter set may be obtained as follows [9]:

- The offset O can be found empirically by inspecting the interferogram on the display by selecting a start point just outside the strong features of the centerburst. Depending on the type of beam splitter and on electronic filtering, the optimum offset may be between 15 and 20 points.
- The optimum number of basis vectors depends on the spectrometer's stability and should be found between 10 and 20.
- The optimum number of points per vector N depends on the spectra of the investigated components and is expected to be between 80 and 200.

As the computation time increases roughly proportionally to the product of M and N , it may be advisable to use values smaller than the optimum ones if very high time resolution is necessary. Although our short overviews of the various kinds of computations to be performed by the minicomputer of an FT-IR spectrometer is, of course, far from complete, we shall now end this series with some concluding remarks about the required performance of the data system.

6 Typical Data System Requirements

The development of fast scanning interferometers in connection with HRGC and the availability of extensive libraries have increased the demand for speed and storage capacity enormously in recent years.

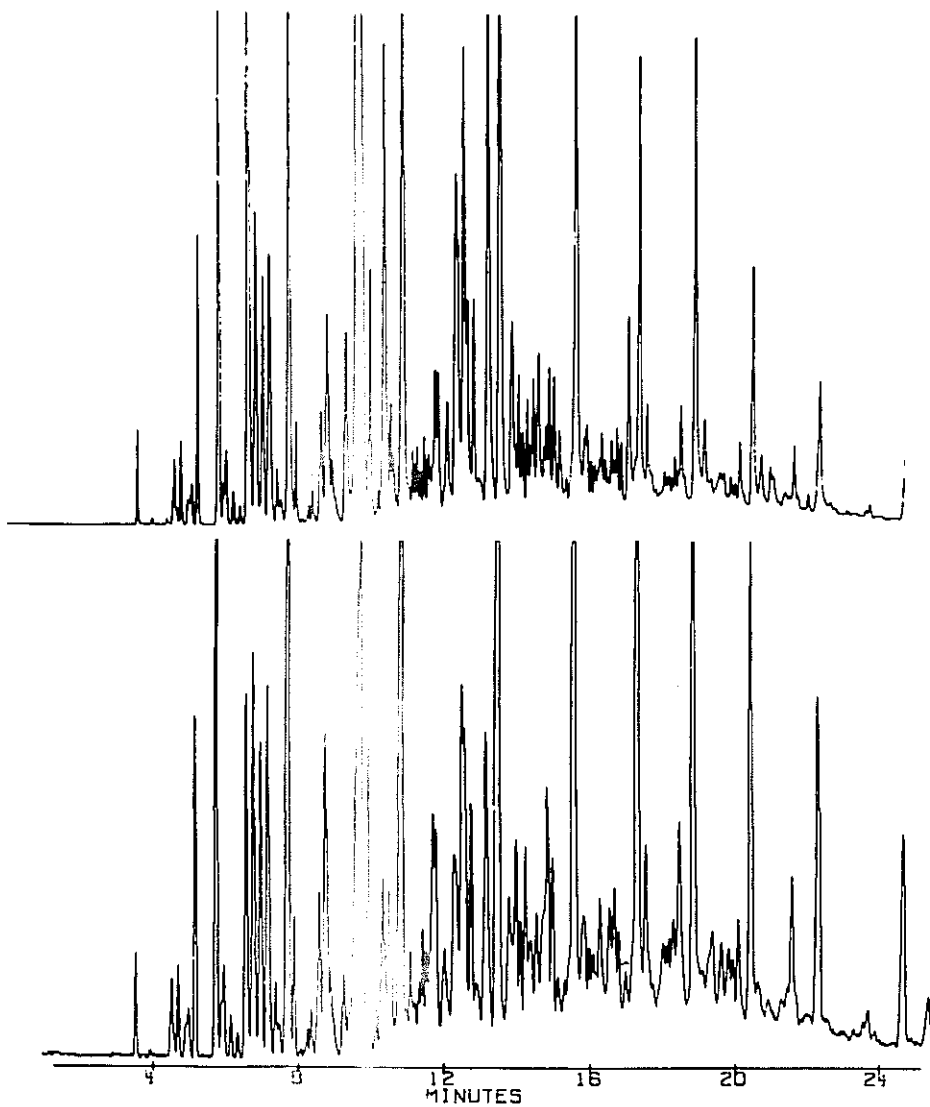
To get an idea of the speed and storage requirements, we consider a HRGC run with 10 - 20 scans per second at a resolution of 8 cm^{-1} . Each scan consists of 2048 data points of 16 bit each. The incoming raw data are written into the computers RAM by DMA either in automatic hardware coadd mode (get current content of a computer word, add

new data, write result back, advance memory address) or in simple replace mode at a rate of about 100 KHz. If every single scan is to be stored on disk, this corresponds to an average disk transfer rate of 20 - 40 Kwords/second (≈ 60 - 120 Kbytes/second with 24 bits/word). A disk of 20 Mbyte capacity would be full after 5 - 10 minutes of GC run. During all this DMA-activity a Gram-Schmidt and possibly several window traces must be computed and displayed. In case of window traces a complete spectrum calculation must also be performed.

In order to meet the speed requirements, not only is sufficiently fast hardware needed but also -and even more important - an operating system with high disk throughput and real time capability. It should be noted that most commercially available operating systems for personal computers would not be suited for HRGC because their file organization is inadequate (no contiguous files), which drastically reduces the transfer rate. In connection with the necessary dedicated hardware this explains why the mini-computer of a FT-IR spectrometer is more than just an 'over-expensive personal computer'.

The required disk space also seems to be ever increasing, especially if large spectral libraries of several tens of thousand spectra are to be stored on a disk, which is also used for GC-IR. Therefore, if optical disks with capacities of several gigabytes become reasonably priced in the next few years, their use might well be interesting.

Modern FT-IR scanners already allow for more than 50 scans per second and new 16-bit ADC's permit sampling rates of 500 KHz, which calls for the development of faster hardware and software. The future will show where these developments will lead.



References

- [1] W. Herres and J. Gronholz, *Comp. Appl. Lab.* 4 (1984), 216.
- [2] J. Gronholz and W. Herres, *Instruments & Computers* 3 (1985), 10.
- [3] M. Born and E. Wolf, "Principles of Solids", Pergamon, Oxford (1970).
- [4] T. Hirschfeld and A.W. Mantz, *Appl. Spectr.* 30 (1976), 552.
- [5] A. Savitzky and M.J.E. Golay, *Anal. Chem.* 36 (1964), 1627.
- [6] J.R. Kauppinen, D.G. Moffat, H.H. Mantsch, and D.G. Cameron *Appl. Spectr.* 35 (1981), 3.
- [7] D.A. Hanna, G. Hangac, B.A. Hohne, C.W. Small, R.C. Wieboldt, and T.L. Isenhour, *J. Chromatogr. Sci.* 17 (1979), 423.
- [8] P.M. Owens, R. B. Lam, and T. L. Isenhour, *Anal. Chem.* 54 (1982), 2344.
- [9] W. Herres, to be published.
- [10] D.T. Sparks, P.M. Owens, S.S. Williams, P. Wang, and T.L. Isenhour, *Appl. Spectr.* 39 (1985), 10.