Comparison of an Integration Procedure to Fourier Transform and Data Averaging Procedures in Chromatographic Data Analysis

Sir: In a recent publication, we presented an integration procedure for improving the limit of detection (LOD) in chromatographic systems (*1*). The integration procedure, though not a data averaging procedure or a simple data frequency filtering procedure, is a technique that suggests data smoothing has occurred in some way. This aspect of the integration procedure was not previously discussed in sufficient detail, within the context of other data "enhancing" procedures. In light of the "comment" by Phillips (*2*), we present here some of the key differences between integration (*1*) and other common data smoothing procedures, as applied to chromatographic data.

FOURIER TRANSFORM CONCEPTS APPLIED TO INTEGRATION

Fourier transform (FT) concepts are readily available in the literature, and will be the basis of our presentation. For our purposes the general application of FT concepts by Lephardt is quite useful (*3*). From that text, if $f(t)$ is a data array originating in the time domain, then $F(w)$ is a data array in the frequency domain obtained via a suitable FT. Further, a mathematical relationship is presented there, in general form, for the calculation of the nth derivative of the time domain array, by manipulation of the frequency domain array (*3*)

$$\frac{d^n f(t)}{dt^n} \rightleftarrows (iw)^n F(w) \qquad (1)$$

where the symbol $\rightleftarrows$ indicates the reversible nature of the FT, t is time, w is frequency, and $i = -1^{1/2}$.

For a typical chromatographic system with a detection time constant and data collection interval of 1 s, all frequency information within a chromatogram is contained between 0 and 1 Hz. The FT (into the frequency domain) of "white" noise, such as that used in the previous paper (*1*), produces a distribution of data, uniform in amplitude and frequency, between 0 and 1 Hz. For a chromatographic peak, without noise, the FT into the frequency domain produces a spectrum that contains most of the peak information at the lower frequency end. The extreme limit of this result is that for an infinitely wide peak in the time domain, the FT into the frequency domain is defined by only the zero point value. An offset base line in the time domain can be thought of as an infinitely wide peak.

It is interesting to describe the integration procedure (*1*) in the context of eq 1. Collecting the entire chromatogram prior to any integration allows for an objective base line adjustment (BLA) procedure, as described earlier (*1*). Any base line offset or long-term drift in the time domain will be observed as extremely low-frequency data in the frequency domain via a FT. The BLA procedure, prior to integration, effectively reduces much of this kind of "noise" in the original chromatographic data array and is equivalent to introducing a very low frequency cutoff in the FT frequency spectrum. In principle, BLA can also be achieved by an analog high-pass filter with a frequency much lower than that of any chromatographic event.

An integrated chromatogram is obtained upon applying running-total integration to the BLA data array (*1*). According to eq 1 this operation is equivalent to substituting $n = -1$. An important observation is that integration in the time domain manifests itself as division by iw in the frequency domain, as correctly pointed out in ref *2*. Thus, higher frequency components of the data are attenuated, while lower frequency components are accentuated. Since most of the original peak information is at lower frequencies, and noise is spread throughout the frequency domain, the implication of eq 1 for integration is enhanced detectability in the integrated chromatogram relative to the original chromatogram. Another feature is the use of the whole peak, or peak area, instead of just the peak height for determining detectability. In terms of eq 1, the shape of an integrated peak, as compared to the original peak, produces frequency domain data that favor the former in S/N, the signal-to-noise ratio. This concept has seemingly been overlooked, or at least not effectively applied in chromatography. The observation that integrated data provide better LOD values as compared to peak height data in graphite furnace/atomic absorbance work (*4*) means that other fields of study have incorporated this point. Yet, our integration procedure (*1*) went further and suggested that detection is possible even if the original data do not provide peak heights that are above the "detection limit". We also suggest that the temporal information for a series of events (chromatographic resolution) need not be degraded while noise is reduced.

GENERAL COMPARISON OF TECHNIQUES

A comparison can be made between the integration procedure proposed earlier (*1*) and both a FT procedure and a data averaging procedure. The FT and data averaging procedures are outlined in Lephardt's chapter (*3*). This comparison is made for procedures that do not significantly broaden chromatographic peak widths by loss of pertinent frequency information within the data. Our integration procedure inherently provides integrated signals that are identical in width as the original chromatographic data. This is obvious in eq 1 with $n = -1$. Even though the high-frequency components are attenuated by iw, they are never discarded. One can always get back the original chromatogram from the integrated chromatogram by doing a formal differentiation (eq 1 with $n = 1$). The exact same number of data points describe the chromatographic event before or after integration, so no loss in *real* chromatographic resolution results. There is, however, a loss in *apparent* resolution as pointed out by ref 2. This is because our visual perception is better adapted to distinguish large changes in slopes (differentiated chromatograms) vs. small changes in slopes (integrated chromatograms). The important point is that resolution also depends on S/N. While differentiated (*5*) or unintegrated (*1*) chromatograms emphasize the inflection points more, they are also more noisy, as pointed out in ref 2, compared to integrated chromatograms.

A FT procedure based upon truncation in the frequency domain provides smoothing that translates into increased S/N upon taking the inverse FT back to the time domain (*3*). In this procedure, knowledge of chromatographic peak frequency components allows one to truncate the FT in the frequency domain at frequencies above the point where the peaks no longer contribute. For this method, there is a trade-off in the process of improving the S/N. With a lowering of the frequency of truncation, integrity (i.e., peak width, resolution, and height information) of the chromatographic data will decrease while the S/N increases. We note that the high-frequency components are lost forever, and indeed there is a real loss of resolution.

0003-2700/86/0358-2093$01.50/0

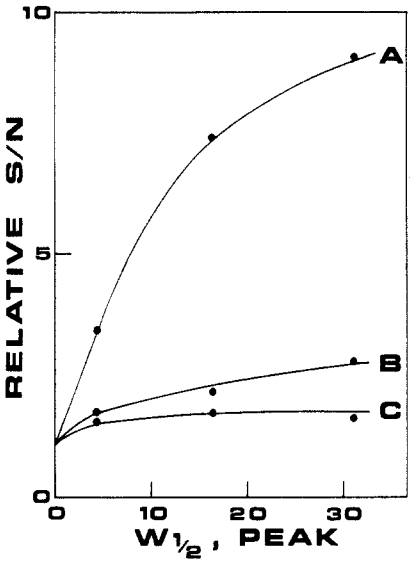

Figure 1. S/N relative to the original data as a function of chromatographic peak width at half height, $W_{1/2}$: A, integration procedure; B, FT procedure; C, data averaging procedure.

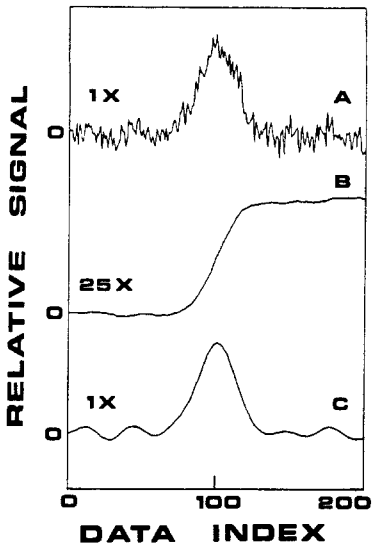

Figure 2. Comparison of the original data (A), applying integration procedure to A (B), and applying FT-truncation procedure to A (C). Scale for each curve is relative and is designated in terms of ×. So, the height of B is about 25× the height of A.

A data averaging procedure is essentially taking the original data and applying a running-average (or a time constant) to the data. For data limited by "white" noise, the theoretical improvement in S/N by data averaging is $N^{1/2}$ were N is the number of points used to calculate the average value at a given point in the chromatogram. Thus, if the data were originally collected with a 1-s time constant, the data averaging procedure effectively provides a N-s time constant. It is anticipated that data averaging will improve the S/N and ultimately the LOD at the expense of broadening the chromatographic peak data. This is because once again the high-frequency components are lost forever. Also, the number of significant points is reduced by $1/N$.

The chromatographic detector noise and peak simulation utilized in ref 1 provided data that were subjected to each of the three procedures outlined. Upon treatment of a given data array the signal and the resulting noise must be measured for each procedure. For the data averaging and FT procedure, the signal is the peak height. The signal for the integration procedure is determined, as in ref 1, as the height of the inflection in the integrated data array. The noise is determined statistically as 5σ noise. The S/N value determined from the original data array acts as a reference point for the three procedures. Laeven and Smit found that integration of noise produces increased uncertainty in quantitation with increased integration time (*6*). Effectively, the larger the number of points that must be integrated to define a peak, the larger the uncertainty in the quantitative results. This suggests that the use of a fixed "event" width, as used earlier (*1*), is not adequate in providing the noise value for the integrated base line. One can instead use a slowly increasing "event" width to define the LOD for the integrated data file. This slowly increasing "event" width is similar in essence as the idea of slowly increasing the time constant applied to the data in the original chromatogram (*7*). For the results obtained, using Gaussian peaks, the event width is equal to $2.55W_{1/2}$, where $W_{1/2}$ is the width of the peak at half height. This is equivalent to $\pm 3\sigma_p$, were σ_p is the peak standard deviation. Accordingly, σ_p is linearly related to retention by (*8*)

$$\sigma_p = \left(\frac{t_0^2}{r}[k(1 + k)] \right)^{1/2} \qquad (2)$$

where t_0 is the dead time, r is the column efficiency (a constant for similar analytes), and k is the solute capacity factor, which is defined in the conventional way. By varying the $W_{1/2}$ for the peak data, we calculated the relative S/N values for the three procedures (integration, FT truncation, data averaging). The results of these calculations are shown in Figure 1. Note that the data averaging procedure was facilitated using N = 4 (4-s time constant). For visual reference in Figure 2, the results of applying the integration and FT procedures can be compared to the original data for a peak with $W_{1/2}$ = 31.1 data units. Some general trends concerning Figures 1 and 2 can be made. The data averaging procedure does not provide a full factor of 2 improvement in S/N as suggested by theory for a four-point average. This is due to loss of peak height in the averaging process. FT followed by truncation of frequency components that do not contain signal information, and subsequent inverse FT, produces an improved S/N as compared to data averaging. The concern with FT "filtering" procedures is in throwing away frequency information pertaining to the analytical signal. The data in Figures 1 and 2 for the FT procedure were calculated at the limit before peak distortion occurs due to over-filtering in the frequency domain, thus obtaining the best S/N possible without severe peak distortion. The integration procedure was applied as reported earlier (*1*). For chromatographic detection systems limited to a great extent by "white" noise, it is clear both graphically (Figure 1) and visually (Figure 2) that the integration procedure provides the greatest improvement in S/N relative to the original data array.

Application of the integration procedure would allow one to quantitate unresolved peaks that may be impossible to quantitate, or even to "detect", in the original data array. Once quantitated, the signal width in the integrated time domain can be compared via eq 2 to diagnose, for a given chromatographic system, the presence of peak overlap and to what extent. Thus, the integration procedure may provide the means to quantitate unresolved peaks that could not be quantitated initially in the original data.

Up to this point, the discussion has dealt with simulated noise and peak data. The assumption was made that chromatographic detector noise is often randomly distributed, in time, about a mean value, such as with "white" noise. An example of real data from an ion chromatography separation followed by UV absorbance detection was studied with the integration procedure. The chromatogram studied recently appeared in the literature (right side of Figure 2, page 59) (*9*). Applying the integration procedure to this chromatogram

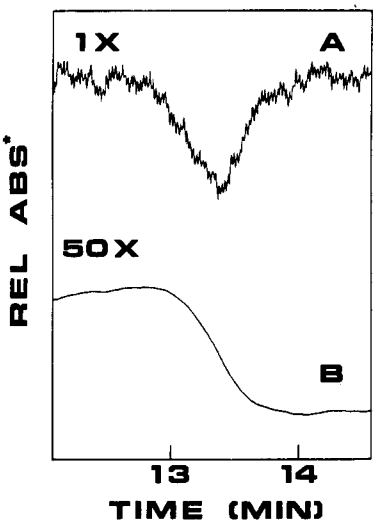

Figure 3. (A) Section of original data from Figure 2, page 59 (*9*). (B) Same section of data as in A after integration procedure was applied to the entire chromatogram. Scale for each is in terms of X. So, the height of B is about 50X the height of A. The peak in A is the ion chromatography separation/UV absorbance detection of SO_4^{2-} from 20 μL injected amount of a 1.2×10^{-3} N solution, using 1×10^{-3} M potassium citrate eluent.

produced an integrated noise value 4.3 times that of the noise in the original chromatogram for an event width of 120 data collection units. This is roughly 1.5 times larger than the value obtained in the simulations (*1*), suggesting that the real-life detector is not exhibiting "white" noise exclusively, but contains some long-term drift components. However, the factor of 1.5 also suggests that white noise is not a bad approximation to the real noise. Figure 3 contains both the original and integrated data for the peak in the chromatogram that elutes at approximately 13.3 min. An improvement in the LOD of 11.6 was determined by establishing confidence limits in both the original and integrated time domains, and measuring the peak signal in each. In the context of absorbance detection in chromatography this corresponds to an extension in absorbance detectability from 2×10^{-4} AU to 1.7×10^{-5} AU. The integration procedure may also be quite useful when coupled with other commercially available detectors. One expects that fluorescence detectors will behave quite similarly as absorption detectors. However, refractive index detectors are much more sensitive to changes in temperature, pressure, solvent composition, etc., so that white noise may not be a good approximation. But then all data smoothing routines will fail when noise has frequency components similar to those of the signal.

SUMMARY

1. The integration method is fully equivalent to the FT representation (eq 2 of ref 2). However, since integration can be done in real time by summation (running total), there is *no need* to perform a forward and then a reverse FT after the whole chromatogram is collected. Besides, the complex reverse transform may present some technical problems.

2. The integration method is actually *more* effective than most frequency-based filters, as shown by Figures 1 and 2 here. The reason is that one can only compare those frequency filters that preserve the integrity of the chromagrographic event. The $1/(iw)$ weighting in integration accentuates the signal at low frequency while gradually suppressing noise at high frequencies. The net effect is to attenuate intermediate frequency noise more strongly. There is a net gain in S/N compared to cutting the high frequencies off completely without altering the low-frequency information.

3. The base line model used is quite realistic, as shown for the real data in Figure 3. Naturally, other detectors may show different noise distributions, limiting the utility of the integration procedure.

4. The value of IMP (*1*) *is* independent of any data averaging that might be applied. This implies that the integration procedure can be used *after* other data averaging techniques have been applied to reduce noise. Further gains in S/N can be expected because the signal is enhanced further by the $1/(iw)$ weighting factor for the remaining frequencies.

5. While the *apparent* chromatographic resolution is degraded because of the final form of the data, the *real* chromatographic resolution in the original data set is preserved by integration. This is because the exact same number of data points (width) represent the chromatographic event before and after integration, and no high-frequency component is discarded completely. To a computer, the confidence level is invariant for identifying an inflection point, a maximum (first derivative), or a zero crossing (second derivative).

6. For cases where standard chromatographic software will not even recognize the existence of "peaks" because of poor S/N, the integration procedure may still be able to provide peak recognition and some quantitative information (*1*). After such a recognition, one may then use chromatographic information (σ_p) to refine the quantitation by defining the limits of integration and to test for unresolved peaks.

LITERATURE CITED

(1) Synovec, R. E.; Yeung, E. S. *Anal. Chem.* **1985**, *57*, 2162–2167.
(2) Phillips, J. B. *Anal. Chem.* **1986**, *58*, 2091–2092.
(3) Lephardt, J. O. In *Transform Techniques in Chemistry*; Griffiths, Peter R., Ed.; Plenum Press: New York; 1978; pp 294–301.
(4) Barnett, W. B.; Bohler, W.; Carnrick, G. R.; Slavin, W. *Spectrochim. Acta, Part B* **1985**, *40B*, 1689–1703.
(5) Woodruff, S. D.; Yeung, E. S. *J. Chromatogr.* **1983**, *260*, 363–369.
(6) Laeven, J. M.; Smit, H. C. *Anal. Chim. Acta* **1985**, *176*, 77–104.
(7) Saadat, S.; Terry, S. C. *Am. Lab. (Fairfield, Conn.)* **1984**, *16*(5), 90–101.
(8) Fritz, J. S.; Scott, D. M. *J. Chromatogr.* **1983**, *271*, 193–212.
(9) Wilson, S. A.; Yeung, E. S. *Anal. Chim. Acta* **1984**, *157*, 53–63.

Robert E. Synovec*
Edward S. Yeung
Ames Laboratory—USDOE and Department of Chemistry
Iowa State University
Ames, Iowa 50011

RECEIVED for review April 10, 1986. Accepted May 5, 1986. R.E.S. thanks the Phillips Petroleum Co. for a fellowship. The Ames Laboratory is operated by the U.S. Department of Energy by Iowa State University under Contract No. W-7405-eng-82. This work was supported by the Office of Basic Energy Sciences.