

CHROM. 9468

EFFECTS OF DIGITAL FILTERS ON CHROMATOGRAPHIC SIGNALS

STUART P. CRAM*

Varian Instrument Division, 2700 Mitchell Drive, Walnut Creek, Calif. 94598 (U.S.A.)

STEPHEN N. CHESLER**

Department of Chemistry, University of Florida, Gainesville, Fla. 32601 (U.S.A.)

and

ARTHUR C. BROWN, III

Varian Instrument Division, 2700 Mitchell Drive, Walnut Creek, Calif. 94598 (U.S.A.)

SUMMARY

Digital filters are quantitatively evaluated for their effects on chromatographic digital data. Changes in peak-shape parameters for Gaussian and asymmetrical chromatographic peaks are described by the error levels introduced. Reductions in baseline noise are discussed in terms of reducing the amplitude and the frequency response. An example of the practical considerations and trade-offs in implementing digital filters in a microprocessor-based data system is presented, so as to minimize peak distortion, and enhance peak detection and signal-to-noise ratio.

INTRODUCTION

Numerical data-handling techniques are extensively applied in chromatographic data systems today to enhance effectively the information content, signal-to-noise ratio (S/N), accuracy, and precision of analytical signals. Many of these advanced techniques have been reviewed by Annino¹ and Hieftje²⁻³. However, minimal attention has been given to the quantitative aspects of digital filters since the publication of the landmark paper in the analytical literature by Savitzky and Golay⁴ in 1964. Their paper treated numerical smoothing and differentiation by least-squares procedures and presented tables of convoluting integers and their normalizing factors. These tables were later corrected⁵, but the effects of the filters and interpretation for their application to chromatographic systems has not been reported. Digital filters are usually applied as part of the signal-processing algorithms of incoming signals to digital data systems and are therefore critical in terms of subsequent data processing and interpretation programs. They are rarely an operator-controlled param-

* To whom correspondence should be addressed.

** Present address: Analytical Chemistry Division, National Bureau of Standards, Washington, D.C. 20234, U.S.A.

eter in chromatography data systems and therefore it is even more important that their effects, potential power, and limitations be understood.

Noise in its various forms is a random error that directly affects the precision and useable range of analytical measurements. Thus it is desirable to remove as much noise as possible without distortion of the signal beyond acceptable limits. Several analog and digital filtering techniques are available.

Passive analog filters (LRC types) have the advantage of low cost and simplicity of design and construction. However, they are (1) limited in most cases to real-time signals, (2) cannot produce derivatives of the input signal (although other analog techniques are available for doing so), (3) must be voltage and impedance matched to the signal source, (4) show variations from one unit to the next owing to component tolerances, and (5) do not lend themselves to systems requiring dynamic changes in filtering characteristics. Active analog filters have the same basic characteristics as passive filters, except that the filtering characteristics can be varied over limited ranges and lower frequencies are more easily filtered. These filters can add amplifier noise to the signal and are complex and expensive to implement. These types of filters are commercially available today for chromatography systems.

Digital filters are numerical counterparts to analog filters. Hard-wired digital filters (1) can implement sophisticated algorithms, such as Fast Fourier Transforms, (2) are very fast, (3) can operate on either real-time or data stored in memory, and (4) can produce derivatives of the input signal. These types of filters are costly, may be complex to design into a system, have limited adaptability, and require accurate digitization.

Software-based digital filters take advantage of the logical and arithmetic capabilities of digital computers and may be adaptive. They (1) are the most versatile in that very sophisticated filters and numerical techniques can be designed, (2) may be changed in real-time and optimized for the S/N and frequency response desired, (3) are simple to implement and use, and (4) allow the versatility of post analysis recalculations. Hard-wired analog and digital filters must be replicated for each input channel to be filtered, whereas any number of different software filters (from 1 to n) can be applied to m channels of incoming signals.

It should be pointed out that, in addition to analog and digital filters, other techniques may also be used for noise reduction or S/N enhancement. These techniques include integration (analog and numerical), modulation, digitization, data bunching, correlation, synchronous detection, multiple pass filters, and averaging. In all cases, these techniques should be characterized in terms of their gain in S/N , frequency response, and effects on precision, resolution, and accuracy (*i.e.*, reduction in systematic error). As an example, it is imperative that the data-bunching rate in chromatography data systems and integrators be set or selectable in terms of the frequency of response (peak width) rather than by the number of data points. The effects of chromatographic noise have been treated in terms of their effect on the precision and accuracy of statistical moments⁶, digitization errors^{7,8}, retention time measurements⁹, detection limits¹⁰, and the design of chromatography data systems to handle sensitive peak detection and spike trapping¹¹.

This paper follows directly from the one by Savitzky and Golay⁴, which is recommended as an introduction to the subject. Therefore, we will treat only moving average (MA) and polynomial smoothing functions as digital filters here in detail.

CHARACTERIZATION OF DIGITAL FILTERS

Digital filters basically serve to attenuate random noise and/or to differentiate the signal. The filters discussed are of the symmetric, non-recursive, linear type which operate only on the input signal. One requirement of these filters is that the digitized data must be in fixed time increments in order to be smoothed with the Savitzky-Golay convolutes⁴. Secondly, data to be filtered will be assumed to be a continuous function such as chromatographic peaks, as either resolved or unresolved elution profiles. High-frequency spikes are "trapped" by such filters and are thus reduced in amplitude. Repetitive spiking at high frequencies may introduce a systematic error in the data, but this is assumed to be a minimal effect in stable, clean chromatographic systems and therefore is not considered in this work. Thirdly, polynomial smoothing requires that an odd number of data points be fit.

Polynomial smoothing functions will be given by

$$Y_p^{(n)} = \sum_{i=-m}^{+m} w_{pi}^{(n)} y_i \quad (1)$$

where

- $Y_p^{(n)}$ = amplitude of the smoothed value
- p = position of smoothed value within convolution interval. $p = 0$ (midpoint) for symmetrical filters
- n = order of the smoothing formula
- $2m + 1$ = number of data points in the smooth
- $w_{pi}^{(n)}$ = convolute coefficient of the i th data point for smoothing operations
- y_i = amplitude of the i th data point before smoothing

such that

$$\sum_{i=-m}^{+m} w_{pi}^{(n)} = 1 \quad (2)$$

and

$$w_{pi}^{(n)} = w_{(-p)(-i)}^{(n)} \quad (3)$$

It should be noted for later reference that if n is even, e.g., $n = 2q$, then

$$w_{pi}^{(2q)} = w_{pi}^{(2q+1)} \quad (4)$$

which says that the formula obtained for finding the smoothed value at the midpoint of the interval ($p = 0$) by using an even-degree polynomial is the same as that obtained with the next higher odd-degree polynomial.

For the special case where $n = 1$, as for linear smoothing formulas,

$$w_{0i}^{(1)} = \frac{1}{2m + 1} \quad (5)$$

it can be seen that

$$Y_p^{(1)} = \sum_{i=-m}^{+m} \frac{y_i}{(2m + 1)} \quad (6)$$

which is the MA digital filter.

Effects of smoothing

In chromatography, the error signal or noise may be considered as random at high frequencies whereas the errors at low frequency are more systematic as they approach the bandwidth of the elution profile. Therefore, it would be desirable to design a lowpass filter for error signals at frequencies higher than that of the signal to be observed. From eqn. 1 it may be shown that the results of a polynomial smoothing function may be expressed as a function of the frequency, ω , in terms of a filter function, given by

$$f(j\omega) = w_0 + 2 \sum_{i=1}^m w_i \cos i \omega \Delta t \quad (7)$$

where

$$\bar{y}(j\omega) = f(j\omega) \cdot y(j\omega) \quad (8)$$

and $\bar{y}(j\omega)$ and $y(j\omega)$ are the Laplace transforms of $\bar{Y}(t)$ and $Y(t)$, respectively.

This may be illustrated by considering a 7-point linear smooth which is given by eqn. 5 as

$$w_{0i}^{(1)} = \frac{1}{7} \quad (9)$$

so that the filter function will be

$$f(j\omega) = \frac{1}{7} (1 + 2 \cos \omega \Delta t + 2 \cos 2 \omega \Delta t + 2 \cos 3 \omega \Delta t) \quad (10)$$

which is illustrated in Fig. 1a. It can be seen that for the frequencies where $f(j\omega) = 1$, the original data will be passed by the smoothing function without change. For frequencies where $f(j\omega) < +1$, the original data will be attenuated by the smoothing function. In the case of $f(j\omega) < 0$, the filter function is given by $|f(j\omega)|$. The negative values relate only to phase shifts in the data.

When data points are sampled at data rate intervals of Δt , data components and noise with frequencies of greater than $1/\Delta t$ will not be completely filtered in the smoothed data. This effect can be predicted from Fig. 1 by the cyclic nature of the filter function, as well as the attenuation factor as a function of frequency. Although this smoothing function is a good lowpass filter, different filter functions may be applied as shown in Fig. 1. The 7-point second-order smooth will follow the true peak shape more closely than a linear smooth and therefore distort the peak shape parameters less. However, it can be seen from Fig. 1b that a second-order filter function will be less effective for high frequencies than one of a lower order when $(2m+1)$ is a constant.

When it is desired to remove the higher-frequency noise components in order to enhance the analytical signal information content, or to develop a filter with a lower bandpass, or one with less cyclic character, a number of alternatives are available; all of which may be predicted from their filter function: (1) The data may be pre-filtered with notched analog cutoff filters, as previously discussed. (2) The sampling time interval, Δt , may be increased so that higher frequencies fall in the region where

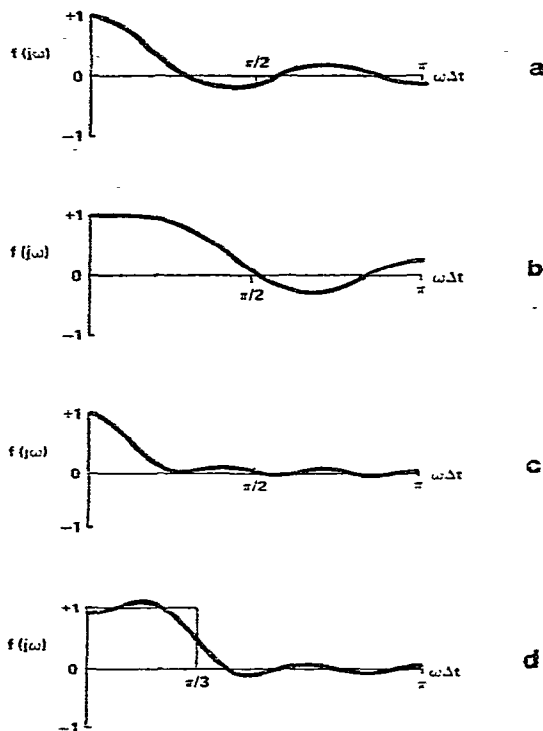


Fig. 1. Filter functions as a function of frequency for (a) a 7-point linear smooth, (b) a 7-point second-order smooth, (c) two passes with a 7-point linear smooth, and (d) a truncated Fourier series.

$\omega\Delta t > \frac{1}{2}\pi$ so that they will be cut off by the analog-to-digital converter (ADC) and further at attenuated by the digital filter. (3) Increase m so as to lower the bandpass of the polynomial filter function. (4) Use a multipass filter as shown in Fig. 7c, where the filter function of k smooths is $[f(j\omega)]^k$. (5) Use operators other than polynomials, as shown in Fig. 7d for a Fourier series. Such functions may be tailored to the functionality of the response curve when it is well defined. All of these filters are extremely versatile, precise, and are well defined when implemented in software. Further it is then possible to update the weighting coefficients on the basis of the incoming signal if it is time dependent, as in the case of chromatography. The results and discussion presented in this work will treat the effects of sampling times, order of the polynomial, and multipass filters (Nos. 2-4 above). Only polynomial and moving average filters will be discussed because they are easier to implement in small data systems and will be shown to be more than adequate for most chromatographic applications when properly applied.

Optimization of the filter band width

Van Rijswick¹² has shown that optimum detection of a Gaussian peak for a filter band width that maximizes the signal-to-noise ratio after filtering, $(S/N)_f$, is given by

$$\left(\frac{S}{N}\right)_f = \frac{0.87A}{\sigma_y w_p} \cdot \left(\frac{w_p}{\Delta t}\right)^{\frac{5}{2}} \left[\frac{K^5}{(1+K^2)^3}\right]^{\frac{1}{2}} \quad (11)$$

where A is the peak area, σ_p the mean amplitude of the normal random noise before filtering, w_p is the Gaussian peak width, Δt is the sampling interval, (note that $w_p/\Delta t$ is the sampling density), and $K = w_f/w_p$ or the ratio of the filter width to the Gaussian peak width. It should be noted that filtering will increase the peak width to $(w_p^2 + w_f^2)^{\frac{1}{2}}$ or $(1 + K^2)^{\frac{1}{2}}$. K will thus reach an optimum at $K = 2.24$ so that for optimum peak detection, the width of the matched filter should be proportional to the width of the chromatographic peak. Although detection limits are not discussed in the results section of this paper, it is important to point out that digital filters may enhance or reduce the detection limits of any chromatographic detector, affect the area recovery and peak width, and must be applied in relationship to the sampling frequency.

EXPERIMENTAL

All of the computer programs for calculations and simulation were run on a PDP-8/L laboratory computer system with 8 K of core (Digital Equipment Corp.). The system includes the following peripheral devices: four-tape magnetic tape cartridge system (Model 4096, Tri-Data Corp.), high-speed paper tape reader (Mark V, Datascan), 15-in. display oscilloscope (Model 1735D, I.T.T.), ASR-33 Teletype (Teletype Corp.), Type 547 laboratory oscilloscope (Tektronix), 10-bit ADC and digital-to-analog converter (Models A811/A618, Digital Equipment Corp.), sample and hold amplifier (A400, Digital Equipment Corp.), multiplexers, a programmable clock, and programmable gain amplifier.

RESULTS AND DISCUSSION

From eqn. 6 it can be seen that the MA filter smooths by taking a running average over $2m + 1$ data points as the convoluting interval. Further, all data points in the smoothing interval are equally weighted and therefore the MA does not track sharp changes in slope. Thus, the most pronounced effect will be seen as a decrease in the peak height, especially for sharp narrow peaks. It should be noted that this means that peak shape parameters will always be distorted by some degree, regardless of the peak shape or symmetry. On the other hand, all symmetrical digital filters treat data points both ahead and behind the data point of interest ($p = 0$), whereas analog filters treat only the current and past data points ($p \leq 0$) and therefore introduce a unidirectional symmetric distortion and error into the results.

Least-squares fits of polynomial filters are more versatile because both the order of the polynomial and the number of data points in the convolute may be varied. Consequently, these functions are better fits for continuous form functions such as chromatographic peaks and introduce less peak distortion than a MA with the same size convolution interval. In terms of high-frequency filtering, the converse is true, as was shown in Fig. 1. If very fast data acquisition rates and small convoluting intervals are used, the polynomial filter will follow the experimental peak profile without significant deviation and effectively filter the noise which is much greater than the bandwidth of the chromatographic peak. Lower-frequency noise will not be filtered as effectively and larger convoluting integrals and/or lower order polynomials will be required.

These effects can be qualitatively predicted by examining the trends in the

normalized convolution coefficients in Fig. 2 for quadratic-cubic polynomial smooths and for quartic-quintic smooths in Fig. 3. These plots show that the coefficients are symmetrical about $p = 0$, and that the greater the number of points in the smooth, the greater the filtering effect will be on both the peak and the noise for a given sampling density. Thus a 7-point smooth will track a changing signal with greater fidelity than will a 19-point smooth. However, the larger smooth will reduce the noise level and frequency more. In compliance with the optimum filter bandwidth criterion, larger smooths require a higher data point density (points per peak width at half-height) to maintain a fixed ratio of filter width to peak width than do smaller smooths. In practice this ratio should be about 1.1 (points in smooth/data point density) for a quadratic-cubic smooth, and about 0.5 for a linear smooth. In the limiting case where the entire peak profile is included within the convolution interval, peak distortion becomes too severe for chromatographic use even with quartic-quintic smooths.

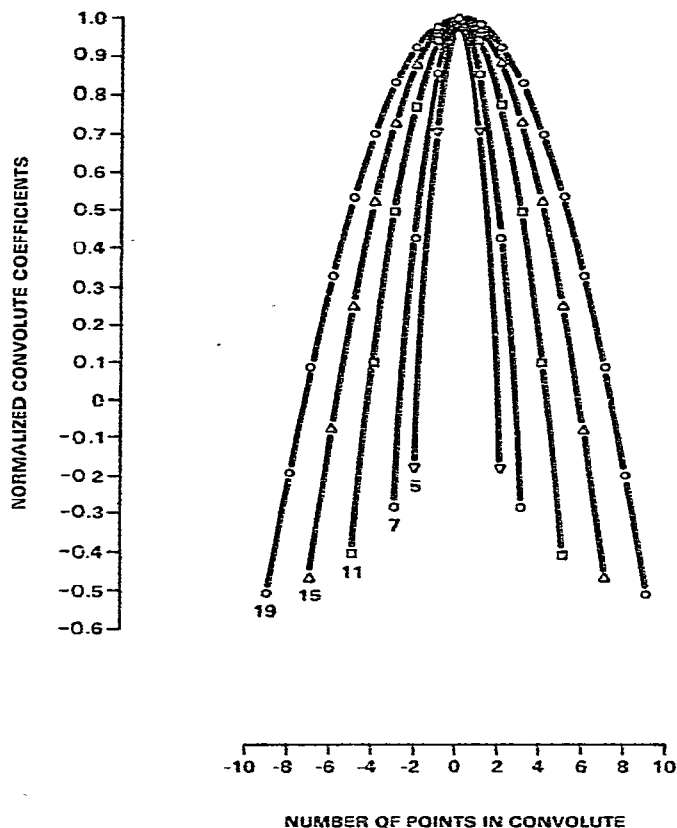


Fig. 2. Convolute coefficients for symmetrical quadratic-cubic smoothing filters plotted about $p = 0$ as a function of p for 5-19-point convolution intervals.

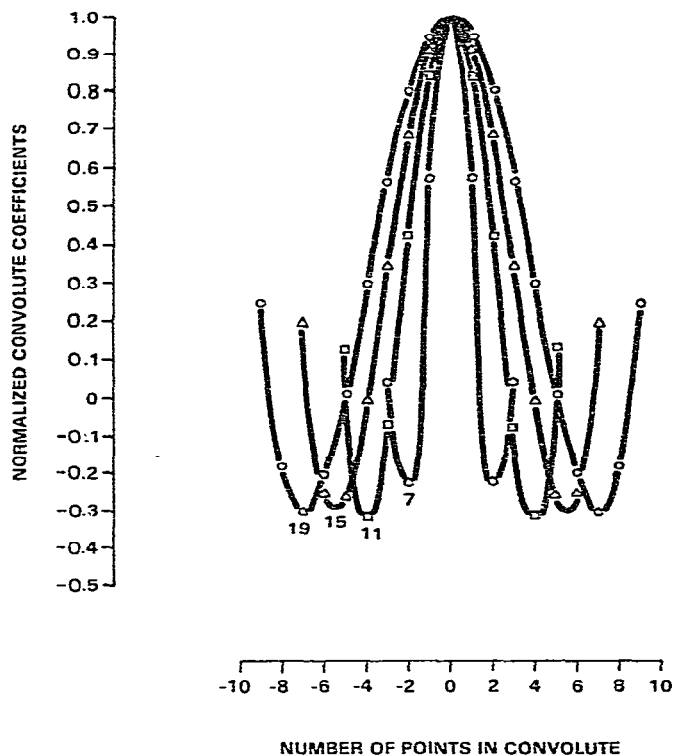


Fig. 3. Convolute coefficients for symmetrical quartic-quintic smoothing filters plotted about $p = 0$ as a function of p for 7-19-point convolution intervals.

Effects on peak shape

Earlier publications described and defined Gaussian and asymmetrical (Type I, II) chromatographic peaks^{6,7}. These peak shapes will be used to illustrate the effects of the digital filters. It should also be noted that a large number of data points were used to represent all peak profiles in order that the accuracy of the peak parameter measurements not be limited by the number of data points per peak⁶. All of the studies reported here were made by computer simulation in order to measure accurately the changes in peak parameters, *i.e.*, the *true* value of all of the peak parameters must be known *before* the errors can be measured.

Peak height was chosen as the simplest peak parameter to illustrate the difference between polynomial and MA smoothing. The relative decrease in the peak height for Gaussian and Type II curves is shown in Fig. 4 for single-pass quartic-quintic polynomial and MA smooths. The asymmetrical peak shows a larger decrease in the peak height than the Gaussian peaks in all cases. The 9-point quartic-quintic smooth decreases the Gaussian peak height by less than 1 part in 10^4 and by 0.38% for the Type II peak so that the peak height ratio is 1.004. In comparison, the 9-point MA smooth decreases the Gaussian peak height by 1.0% and the Type II peak by 5.94% to give a peak height ratio of 1.05. Therefore, as predicted from the discussion earlier in the paper, the MA smooth will always be more detrimental to retaining the

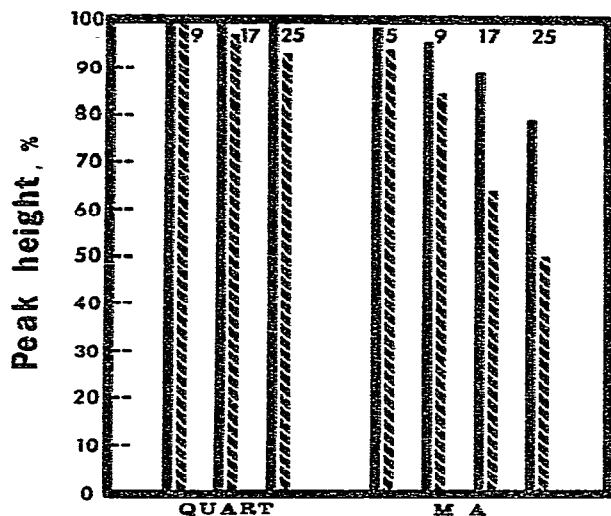


Fig. 4. Plot of the percent of original peak height after filtering with a single smooth as a function of the number of points in the convolute interval for quartic-quintic (QUART) and moving average (MA) smooths. Solid and broken lines represent Gaussian (>75 data points/peak) and Type II (>150 data points/peak) peaks, respectively.

integrity of the peak height because of the sharp change in slope. Further, it may be seen that large convolutes are not justifiable for doing good quantitative computer-based chromatography because the peak height ratio (Gaussian/Type II) decreases from 1.004 to 1.073 in going from a 9- to a 25-point quartic-quintic smooth and from 1.05 to 1.62 over the range of a 5- to a 25-point MA smooth. Since true Gaussian peaks are rarely obtained in GC or liquid chromatographic practice, it is particularly important to realize that, as the peak asymmetry increases, the peak height losses will increase and small convolutes with higher-order polynomial filters become the only method of choice. Type I peaks are not shown as they represent an intermediate case between the Gaussian and Type II peaks.

Because the Type II peak shape is more realistic in experimental work, the effects of the type of filter, number of points in the convolute, and the sampling density ($w_p/\Delta t$) are shown in Fig. 5. It is readily apparent that any digital filter applied to a peak where the sampling density is too low has disastrous effects. A minimum of 3.3 data points/ σ is required for accurate Gaussian peak area characterization according to the criteria of Chesler and Cram⁶. By comparing 9-point convolutes for the higher sampling density it is seen that the quartic-quintic smooth (97.4%) is more accurate (in terms of peak height recovery) than the quadratic-cubic (94.5%) which is more accurate than the MA (79.3%). However, when the same peaks, filters, and convolutes are compared in the frequency domain, the MA is more effective than the quadratic-cubic which is more effective than the quartic-quintic in reducing the peak-to-peak amplitude of normal random noise. Fig. 5 also shows that to recover $\geq 90\%$ of the peak height at a sampling density of 3.3 points/ σ the following filters could be used for a Type II peak: 7-19-point quartic-quintic, 5-13-point quadratic-cubic, or a 3-5-point MA.

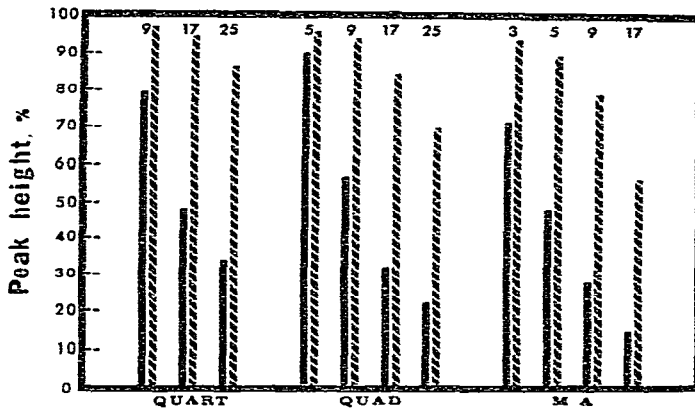


Fig. 5. Plot of the percent of original peak height of a Type II peak after filtering with a single smooth as a function of the number of points in the convolute interval for quartic-quintic (QUART), quadratic-cubic (QUAD), and moving average (MA) smooths. Solid lines represent a sampling density of 1 data point/ σ and broken lines are 3.3 points/ σ .

Fig. 6 is a summary of the trends of the data in Table I. The error in the filtered data for the zeroth moment (u_0 , area) of the Type II peak is consistently larger than for a Gaussian by approximately an order of magnitude, regardless of the type of smooth or the size of the convolute interval. In terms of peak area recovery, it may be seen that only the ≥ 25 -point filters will introduce significant errors in the peak area. This is because of the relative weighting of the peak symmetry and the use of

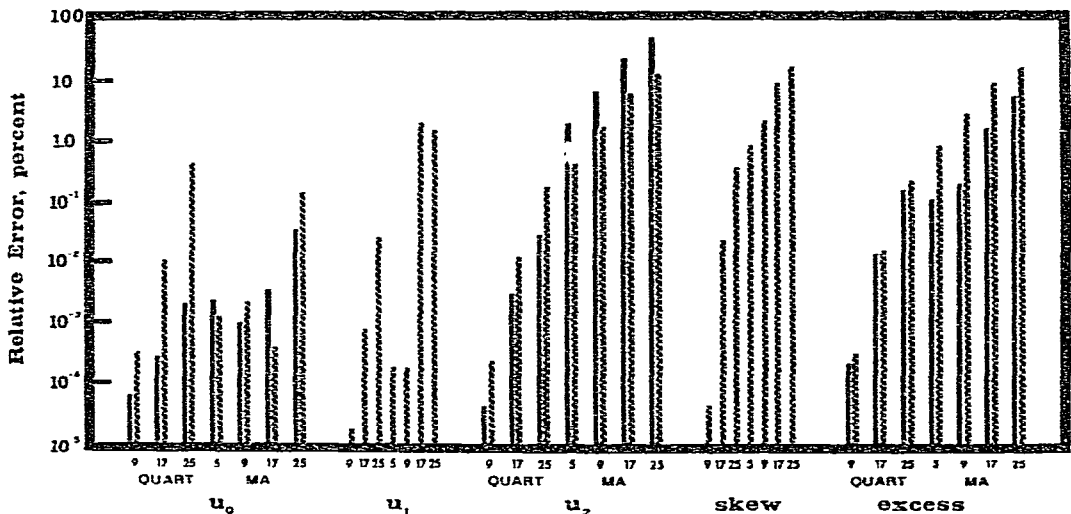


Fig. 6. Plot of the relative error after a single smooth in the zeroth moment (u_0 , peak area), first moment (u_1 , retention time), second moment (u_2 , peak width), skew, and excess for Gaussian (solid lines) and Type II (broken lines) peaks after applying quadratic-cubic (QUAD) and moving average (MA) smooths. The peak integration limits are $\pm 0.1\%$ for sampling densities of ≥ 75 points/peak for the Gaussian and ≥ 150 points/peak for the Type II peak.

TABLE I

RELATIVE ERRORS IN CHROMATOGRAPHIC PEAK SHAPE PARAMETERS AS A FUNCTION OF PEAK SYMMETRY AND FILTER FUNCTION

	u_0	u_1	u_2	u_3	u_4	Skew	Excess
<i>(I) Gaussian peaks (≥ 75 data points/peak)</i>							
(a) No smooth value	1.00	0.00	1.00	0.00	3.00	0.00	3.00
(b) Relative error, %							
9-pt. quadratic-cubic	$6 \cdot 10^{-5}$	—	$4 \cdot 10^{-5}$	—	$3 \cdot 10^{-4}$	—	$2 \cdot 10^{-4}$
17-pt. quadratic-cubic	$3 \cdot 10^{-4}$	—	$3 \cdot 10^{-3}$	—	$2 \cdot 10^{-2}$	—	$1 \cdot 10^{-2}$
25-pt. quadratic-cubic	$2 \cdot 10^{-3}$	—	$3 \cdot 10^{-2}$	—	$2 \cdot 10^{-1}$	—	$2 \cdot 10^{-1}$
5-pt. MA	$2 \cdot 10^{-3}$	—	2.0	—	3.9	—	$1 \cdot 10^{-1}$
9-pt. MA	$1 \cdot 10^{-3}$	—	6.7	—	14	—	$2 \cdot 10^{-1}$
17-pt. MA	$4 \cdot 10^{-3}$	—	24	—	51	—	1.6
25-pt. MA	$4 \cdot 10^{-2}$	—	51	—	117	—	5.4
<i>(II) Type II peaks (≥ 150 data points/peak)</i>							
(a) No smooth value	1.58	2.54	23.3	273	$5.51 \cdot 10^3$	2.42	10.1
(b) Relative error, %							
9-pt. quadratic-cubic	$3 \cdot 10^{-4}$	$2 \cdot 10^{-5}$	$2 \cdot 10^{-4}$	$1 \cdot 10^{-4}$	$2 \cdot 10^{-4}$	$5 \cdot 10^{-2}$	$3 \cdot 10^{-4}$
17-pt. quadratic-cubic	$1 \cdot 10^{-2}$	$8 \cdot 10^{-4}$	$1 \cdot 10^{-2}$	$5 \cdot 10^{-3}$	$9 \cdot 10^{-3}$	$2 \cdot 10^{-2}$	$2 \cdot 10^{-2}$
25-pt. quadratic-cubic	$5 \cdot 10^{-1}$	$3 \cdot 10^{-2}$	$2 \cdot 10^{-1}$	$9 \cdot 10^{-2}$	$1 \cdot 10^{-1}$	$4 \cdot 10^{-1}$	$2 \cdot 10^{-1}$
5-pt. MA	$1 \cdot 10^{-3}$	$2 \cdot 10^{-4}$	$5 \cdot 10^{-1}$	$4 \cdot 10^{-2}$	$2 \cdot 10^{-1}$	$8 \cdot 10^{-1}$	$8 \cdot 10^{-1}$
9-pt. MA	$2 \cdot 10^{-3}$	$2 \cdot 10^{-4}$	1.7	$1 \cdot 10^{-1}$	$8 \cdot 10^{-1}$	2.7	2.6
17-pt. MA	$4 \cdot 10^{-4}$	2.0	6.2	$5 \cdot 10^{-1}$	2.9	9.2	8.8
25-pt. MA	$2 \cdot 10^{-1}$	1.5	13	$4 \cdot 10^{-1}$	7.4	18	17

symmetrical filter functions. For all filters with smaller convolutes, the accuracy of peak area recovery will be limited by the errors associated with sensing peak start and end rather than by the filter. Therefore peak area will always be the least sensitive peak parameter to distortion by digital filters. In all cases, the error in measuring a filtered peak area will increase as the peak asymmetry increases.

The effect of quadratic-cubic and MA smooths on the first moment (u_1 , center of gravity) are negligible for a Gaussian peak. The errors in u_1 increase markedly as the convoluting integral is made larger for a fixed data set. However, even a 25-point quadratic-cubic smooth will only shift the retention time by 0.5 sec for a peak eluting in 30 min while the same peak at the same retention time will increase by 36 sec if a 25-point MA smooth is used.

The error in the second moment (u_2 , variance) was discussed earlier in the paper for Gaussian peaks and the resultant is that the Type II peak will give smaller errors than the Gaussian for an MA smooth. The magnitude of the peak width error for the Gaussian may be derived from eqn. 11. The errors in u_2 in a Type II peak are larger than those for a Gaussian peak using a polynomial smoothing function, although most reasonably sized filters will have a negligible effect. This is important in that the resolution and efficiency of the column are not degraded by the filter function. However, a 9-point MA filter could result in an apparent decrease in column resolution from 1.0 to 0.94 for a Gaussian peak.

The relative error introduced in skew of a Gaussian peak is negligible because the skew is zero by definition. Therefore only the results for the Type II peak are shown in Fig. 6. Skew (and excess) are seen to be particularly sensitive to digital filtering. This is caused by the weighting of the tail of the peak in the calculation of u_3 and the high sensitivity to even minor changes in the profile of the tail. Thus the results are predictable: polynomial smooths will always introduce smaller errors than a MA, and the higher the order of the polynomial and the smaller the convoluting interval, the smaller the peak symmetry distortion.

There is no effect of peak symmetry after smoothing with a polynomial on the excess, which indicates that the relative error introduced in the peak flatness and the increase in peak variance are of the same magnitude. This is not true for a MA smooth because of the effect of equally weighting data points ahead and behind the data point of interest ($p = 0$) with asymmetric peaks. Therefore the effect is to flatten the peak more than to broaden it. In other words, the fourth moment (u_4) is extremely sensitive to shifts in the distribution of the peak area away from the center of gravity. The large errors for the skew and excess, shown in Fig. 6, indicate that meaningful measurements will only be made for these parameters by using low-noise chromatographic systems and peaks with a large signal-to-noise ratio.

The effects of multiple smooths were described earlier as increasing the filter function by the k th power for k consecutive smooths, and decreasing the frequency bandwidth. The effects of ten consecutive smooths are shown in Fig. 7 to show the accuracy of polynomial smooths for all reasonable chromatographic peak shapes and another reason for avoiding MA smoothing. The compounded errors in the Gaussian peak heights are: 0.37% for the quadratic-cubic, $< 10^{-2}\%$ for the quartic-quintic,

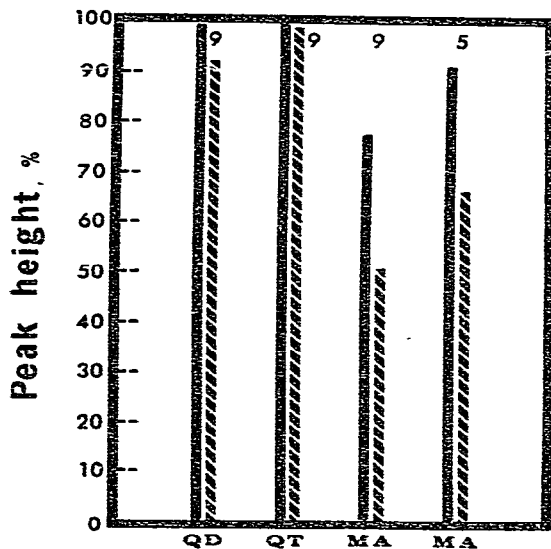


Fig. 7. Plot of the percent of original peak height after ten consecutive smooths as a function of the number of points in the convolute interval for quadratic-cubic (QD), quintic-quartic (QT), and moving average (MA) smooths. Solid and broken lines represent Gaussian (> 85 data points/peak) and Type II (> 210 data points/peak) peaks, respectively.

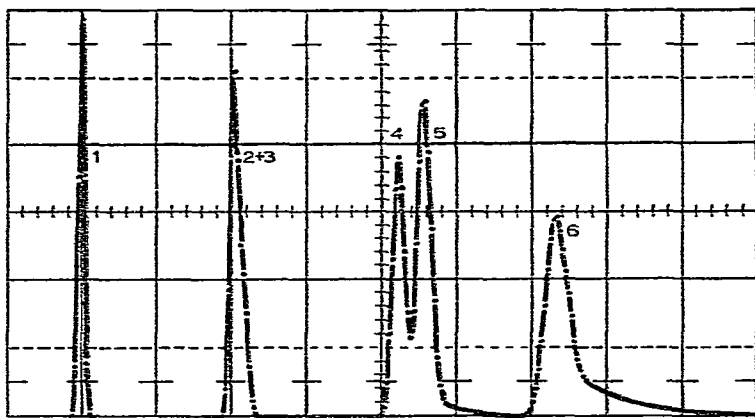


Fig. 8. Simulated 6-component chromatogram shown without smoothing. Peak No. 1 is normalized to 100% of full scale. Peak sampling densities are given in Table II.

22.7% for the 9-point MA, and 8.75% for the 5-point MA. The corresponding errors for the Type II peak are: 7.76, 1.25, 49.6, and 29.6%, respectively.

Figs. 8-14 are simulated chromatograms to show graphically the effects of applying different digital filters. The 6-component chromatograms are simulated to illustrate the effects on a sharp narrow peak (No. 1), a sharp peak with an unresolved shoulder (Nos. 2 and 3), a pair of overlapping peaks (Nos. 4 and 5), and a broad, skewed peak (No. 6). Fig. 8 shows the original, unsmoothed chromatogram, normalized to the peak height of peak No. 1. Table II describes the sampling density for the

TABLE II

PEAK HEIGHTS OF THE SIMULATED CHROMATOGRAM (FIG. 8) MEASURED AFTER DIGITAL FILTERING*

	Peak height (%)	
	Peak No. 1	Peak No. 6
No smoothing	100	100
3-pt. MA	72	94
5-pt. MA	48	90
9-pt. MA	28	79
17-pt. MA	15	56
5-pt. quadratic-cubic	86	96
9-pt. quadratic-cubic	57	94
17-pt. quadratic-cubic	32	84
25-pt. quadratic-cubic	22	70
9-pt. quartic-quintic	80	97
17-pt. quartic-quintic	48	95
25-pt. quartic-quintic	34	86

* Sampling density for the simulated chromatogram (data points/ σ): Peak No. 1 = 1.0; No. 2 = 1.25; No. 3 = 1.43; No. 4 = 2.0; No. 5 = 2.22; No. 6 = 3.33.

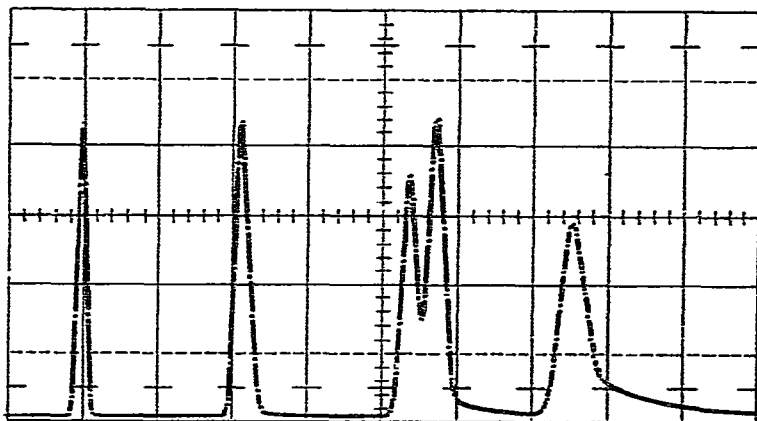


Fig. 9. Simulated 6-component chromatogram after smoothing with a 3-point MA. Scaling factors same as in Fig. 8.

peaks and the attenuation factors for the sharp narrow peak with a typical but insufficient sampling density (No. 1) and the broad peak with an acceptable sampling density (No. 6). Note that in the chromatogram in Fig. 8, the shoulder on peak No. 2 is not discernable in that no peak broadening is apparent, and that the valley between peaks Nos. 4 and 5 is about 25% of the peak height of peak No. 5.

Using the same scaling factors for display as in Fig. 8, the results of 3- and 9-point MA smooths are shown in Figs. 9 and 10. The peak height recoveries, normalized to those in Fig. 8, are reported in Table II. More importantly, note that the 3-point MA has not seriously perturbed the peak shape, resolution, peak tail, or baseline compared with Fig. 8, even though the sampling density is too low on all of the peaks. However, Fig. 10 shows the striking contrast between a 3- and 9-point MA, where the differences are accentuated by the low sampling density. In Fig. 10, peak No. 3 serves to broaden and skew peak No. 2 and peaks Nos. 4 and 5 have been fused by

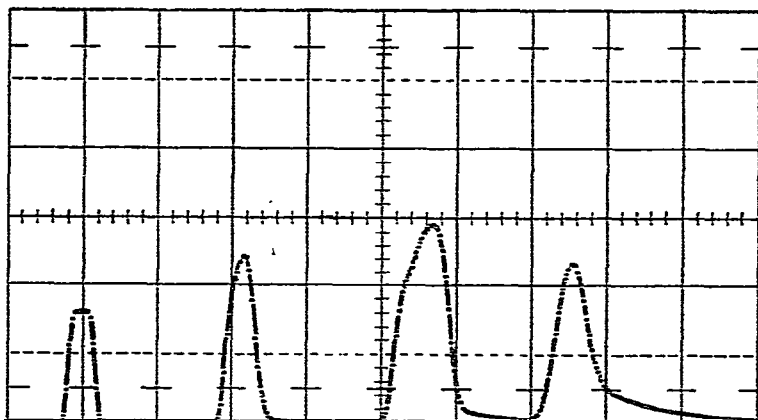


Fig. 10. Simulated 6-component chromatogram after smoothing with a 9-point MA. Scaling factors same as in Fig. 8.

the filter. Similar comparisons may be made between Figs. 11 and 12 for the 9- and 17-point quadratic-cubic and Figs. 13 and 14 for the quartic-quintic smooth of the same convolute intervals. The striking distinction between the 17-point polynomial smooths in Figs. 12 and 14 is the resolution between peaks Nos. 3 and 4 which illustrates the power of the higher-order polynomials as previously discussed. The "ringing" on the leading and trailing edges of the peaks is a consequence of the convoluting coefficients for higher-order polynomials where the sampling density is too low. The extreme case is where the convoluting interval is larger than the number of data points in the peak. In this case "ringing" and peak area "amplification" will result as shown in Fig. 12.

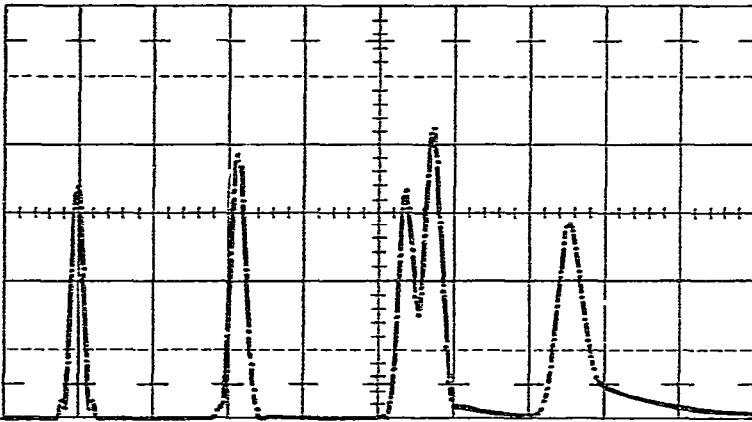


Fig. 11. Simulated 6-component chromatogram after smoothing with a 9-point quadratic-cubic. Scaling factors same as in Fig. 8.

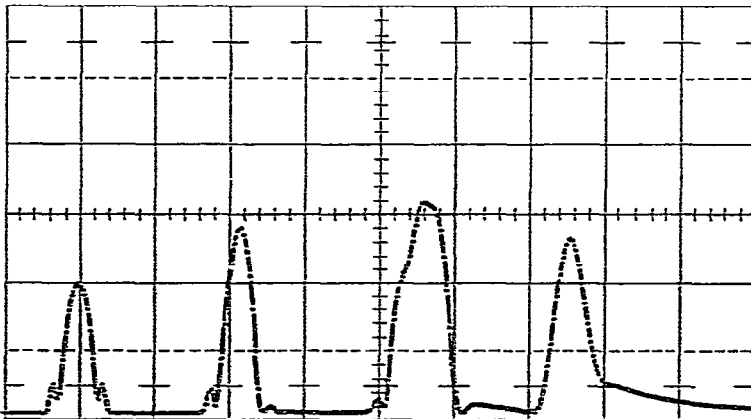


Fig. 12. Simulated 6-component chromatogram after smoothing with a 17-point quadratic-cubic. Scaling factors same as in Fig. 8.

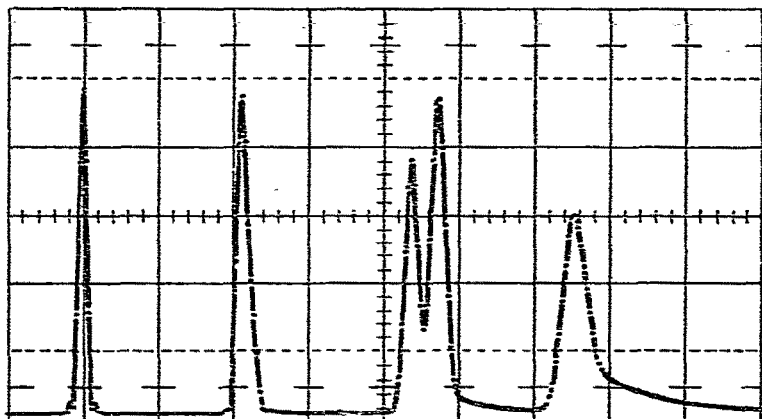


Fig. 13. Simulated 6-component chromatogram after smoothing with a 9-point quartic-quintic. Scaling factors same as in Fig. 8.

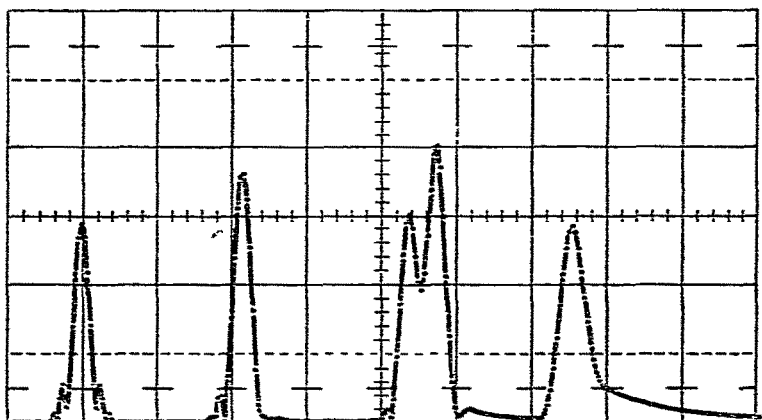


Fig. 14. Simulated 6-component chromatogram after smoothing with a 17-point quartic-quintic. Scaling factors same as in Fig. 8.

Effects on noise

The objective of this paper is to treat the effects of digital filters on chromatographic peaks and peak shape parameters. In addition to peak or signal considerations, we will briefly illustrate the effects on noise, although these effects are well known and have been treated comprehensively elsewhere^{13,14}. To quantitate the noise attenuation, we have defined a "filter factor", Φ , by

$$\Phi = \frac{\sigma_n - \sigma_s}{\sigma_n} \quad (12)$$

where σ_n is the standard deviation of the normal random noise and σ_s is the standard deviation of the noise after smoothing. Thus for $\Phi = 0$ there is no smoothing effect because $\sigma_n = \sigma_s$, and $\Phi = 1$ represents complete smoothing such that the noise is below the detection limit ($\sigma_s = 0$). Figs. 15-17 illustrate the reduction of the normal random baseline noise before and after filtering for the 9-point MA, quadratic-cubic,

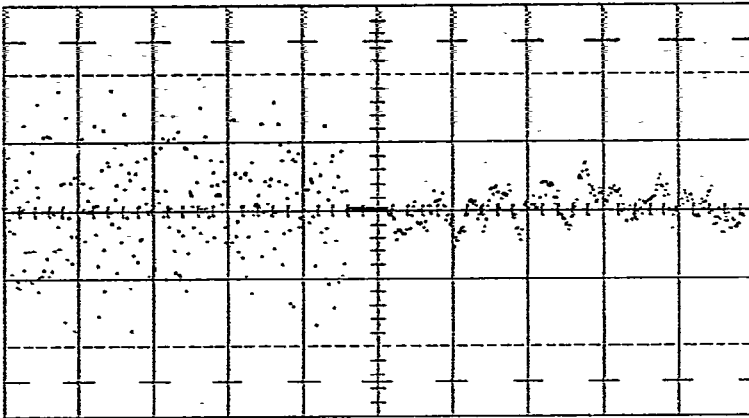


Fig. 15. Simulated normal random noise on a chromatographic baseline. Left-hand side is original noise trace normalized to $2/3$ of full scale. Right-hand side is the same trace, displayed after smoothing with a 9-point MA.

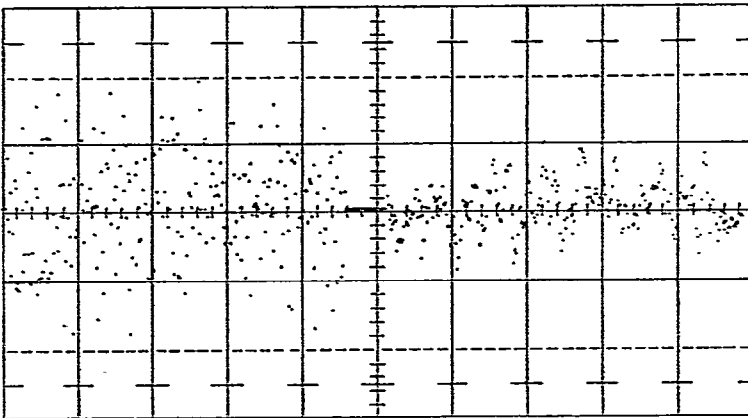


Fig. 16. Simulated normal random noise on a chromatographic baseline. Left-hand side is original noise trace normalized to $2/3$ of full scale. Right-hand side is the same trace, displayed after smoothing with 9-point quadratic-cubic.

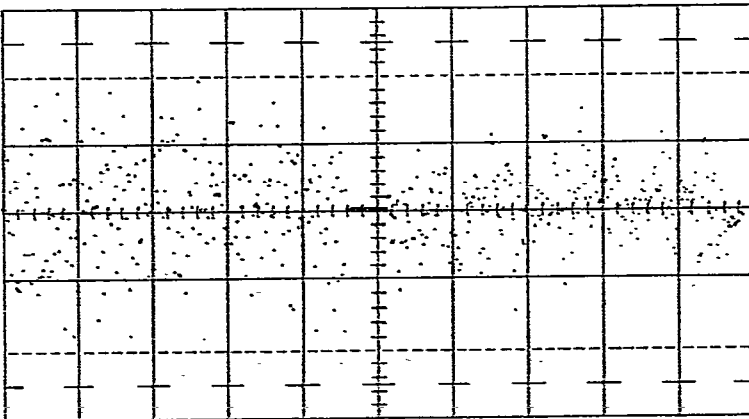


Fig. 17. Simulated normal random noise on a chromatographic baseline. Left-hand side is original noise trace normalized to $2/3$ of full scale. Right-hand side is the same trace, displayed after smoothing with a 9-point quartic-quintic.

and quartic–quintic smooths, respectively. The left-hand side of the baseline segment is the unfiltered noise display, normalized to two-thirds of full scale for display, and the right-hand side of the trace is the same baseline segment after filtering. The filter factors are given in Table III. Where the filter factor is large, the lower-frequency noise components may be observed (*e.g.*, Fig. 15) which is often referred to as “wander” in the chromatographic literature.

TABLE III
ATTENUATION OF NORMAL RANDOM NOISE

Filter function	Filter factor
9-pt. MA	0.702
9-pt. quadratic–cubic	0.553
9-pt. quartic–quintic	0.387
25-pt. MA	0.825
25-pt. quadratic–cubic	0.733
25-pt. quartic–quintic	0.667

Fig. 18 compares the MA and polynomial filters treated in this paper in terms of the filter factor for each type as a function of the number of points in the convolute. Thus, a 5-point MA filter will reduce normal random noise with the same filter factor as an 11-point quadratic–cubic and a 17-point quartic–quintic smooth. As the number of points in the convolute increases, the filter factor becomes less affected by the type or order of the smooth. For the smaller, more commonly used convolute intervals, Fig. 18 shows that the filter factor may range from 0.25 for a 7-point quartic–quintic filter to 0.65 for a 7-point MA. Note that the filter factor changes fastest in this range of convolutes.

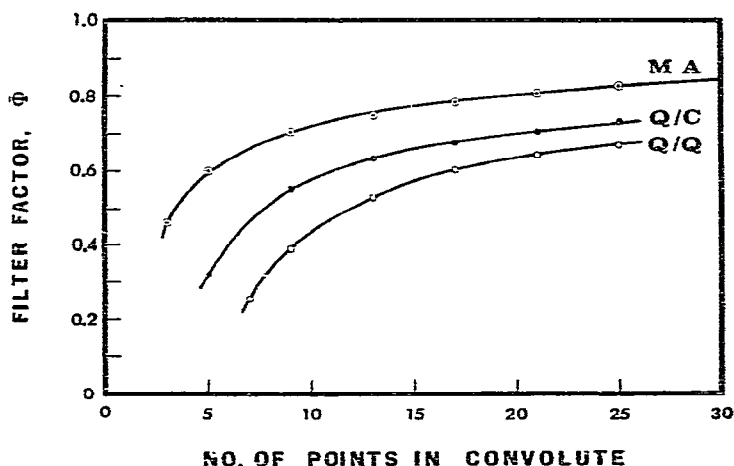


Fig. 18. Response of the filter factor for normal random noise as a function of the number of points in a convolute interval for moving average (MA), quadratic–cubic (Q/C), and quartic–quintic (Q/Q) smoothing functions.

Fig. 19 shows the cutoff filter frequencies for 5-, 17-, and 19-point quadratic-cubic filters in terms of their effectiveness (Φ). The smaller the convolute interval, the higher the bandpass of the filter for a given polynomial order and this is in agreement with the theory. For the range shown in Fig. 19, the midpoint of the frequency cutoff ranges from 20 to 70 Hz which is *not* a large range. Therefore principal consideration should be given to choosing a filter on the basis of the peak parameter to be measured and the amplitude of the noise. This is the experimental verification analogous to the theoretical curves shown in Fig. 1, which emphasizes the fact that a digital filter may be "tuned" to a cutoff frequency by design if the noise has been characterized and the filter factor is known as a function of the frequency, the number of points in the convolute interval, and the order of the polynomial.

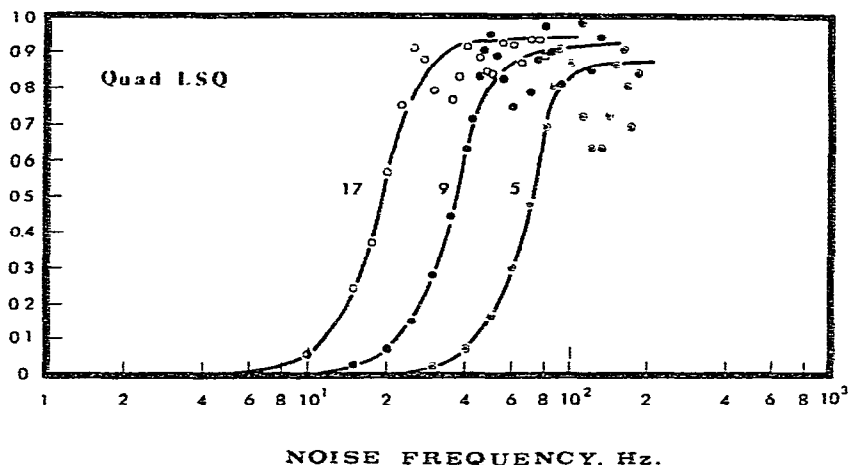


Fig. 19. Frequency response of 5-, 9- and 17-point quadratic-cubic digital filters for normal random noise as measured by the filter factor.

From Figs. 8-15 it can be seen that S/N may be significantly improved to decrease detection limits. At the same time, it is pointed out that peak area, retention time, resolution, and all other peak parameters for real chromatographic peaks are not conserved. It has been shown that all kinds of digital filters may be used to enhance S/N and that the error level is dependent upon: (1) number of data points/ σ ; (2) band width of the peak; (3) amplitude of the noise; (4) frequency distribution of the noise; (5) peak shape parameter to be measured; (6) S/N ; and (7) accuracy desired.

IMPLEMENTATION IN PRACTICE

The Varian CDS-111 and CDS-101 (Varian Instrument Division) are new, dedicated, small chromatography data systems which serve as examples of the practical considerations and trade-offs in designing and implementing digital filters. The following basic objectives governed the design of the filter algorithm in order to make the system as generally applicable, accurate, precise, and reliable as possible within the constraints of a microprocessor-based system. The filter must:

(1) Reduce medium- and low-frequency noise from the column and detector. This noise is close to the frequency of the peak: $F_{\text{noise}} \geq 3 F_{\text{peak}}$. High-frequency noise is attenuated by a two-pole analog filter and an integrating ADC (attenuation of 50- or 60-Hz power line noise and harmonics).

(2) Conserve peak area, and limit the distortion of peak height and width. The peak area measurements must not be adversely affected by the digital filter (*i.e.*, no side lobes or generation of false peaks, no distortion of peak shapes to affect area measurement on fused peaks).

(3) Calculate the first derivative of the detector signal (slope) for use in peak detection (peak-baseline discrimination). Reduce the slope noise due to the "frequency gain" characteristic of the differentiation process. The slope values must be reliable (accurate, precise, and continuous over the full domain of values) so that the peak detection programs do not have to contain the more complex noise-peak-baseline discrimination routines. Slope values for narrow peaks must cover a wide range while slope values for wide peaks must have high resolution.

(4) Operate with an existing ADC producing 20 conversions per second, with a resolution of $1 \mu\text{V}$ at the most sensitive range, and autoranging to 4, 32 and 256 μV at higher voltage levels. Total dynamic range is 10^6 . The ADC translates a smoothly changing signal into a stair-step progression of digital values. With slowly changing signals (*i.e.*, baseline), the ADC values remain constant until the signal moves into the next digital value. This produces a spike or pulse in the first derivative, which should be suppressed to the extent of not affecting the peak detection programs. This effect is not present when the noise level of the signal equals or exceeds the resolution of the ADC.

(5) Be self-adapting to changes in peak frequency, specifically to self-update filtering when peak widths increase. Also to adjust simultaneously the peak detection threshold level (the slope) to take advantage of the reduction of noise due to the filtering increase. Fixed-frequency filters are not suitable to isothermal or isocratic chromatographic separations owing to the increase in peak widths. The filter should not self-adjust when peak widths remain constant as in temperature or gradient programmed separations. The operator should be able to change the filter for different peak widths at will.

(6) Be implemented in an 8-bit, binary microcomputer having an average instruction execution time of $30 \mu\text{sec}$, and operate at the ADC speed (20 data points per second).

The digital filter includes four software elements: a variable pre-integration step to provide the variable frequency response characteristic (also known as data bunching); a 4-point symmetrical MA to suppress momentary signal curvature due to noise, but not suppress the trend of the slope caused by a real peak; a 5-point quadratic least-squares estimate of the first derivative; and an algorithm for the self-adjusting portion of the filter.

The pre-integration step is disabled at the minimum peak width setting, leaving only the MA and first derivative calculation in operation. At wider peak widths, data points are summed in groups of 2, 4, 8... up to 256. This causes two effects: first, the time interval between the 4 data points in the MA and the 5 data points in the slope calculation is increased. The time interval changes from $1/20$ sec with no pre-integration to $1/10$ sec with 2 points summed, and so on. This spreading effect

produces a reduction in noise by $\sqrt{2}$ for each doubling of the number of data points summed. The second effect is the reduction of noise produced by the integration process itself. This also produces noise reduction of $\sqrt{2}$, and thus the combined noise reduction is $(\sqrt{2})^2$ or 2. The dynamic range of the filter frequency is 256:1 using this technique.

The MA smooth was selected because of its high efficiency in suppressing curvature (change in slope) of the data points contained within the MA. Because of its potential for severe shape distortion, a small MA was selected. This allows efficient piece-wise smoothing of the peak profile while still maintaining the peak shape. A 4-point average was selected instead of the more usual 3- or 5-point averages because of an area integration consideration. The 4-point average only includes 20% of the peak width at half-height at any time (20 data points per peak width at half-height are used, or 8 data points/ σ), which minimizes the peak shape distortion. The 4-point average produces a smoothed value at the midpoint of the average where in fact no data point exists. This $\frac{1}{2}$ -data point offset is used by the area integration routine to help compensate for the fact that valley points are located $\frac{1}{2}$ data point late on the average. This improves the accuracy of peak area measurement on fused peaks if the results of replicate analyses are averaged.

The 5-point slope calculation uses the convolution technique. Here the least-square estimate is calculated based on the data points after going through pre-integration and the MA. One effect of pre-integration is scaling of the first derivative values. As the number of points in the pre-integration step increases, the slope values increase as the square of the increase. This is due in part to the design of the filtering program because all normalization factors are grouped together at one operation. However, the scaling of the slope is reduced to a linear relationship of the number of points in the pre-integration step, and this produces the exact increase in peak detection sensitivity that can be made owing to the decrease in noise produced by the change in the number of points used in the pre-integration step.

Finally the filter is self-adapting by responding to the peak width of the currently eluting chromatographic peaks. When a peak is 50% wider than the expected peak width, the filter doubles the number of points in the pre-integration step. The change takes place only at valley points or on the baseline following the peak. To reduce the effect of peak tailing the time from peak start to the peak maximum is used as the measure of peak width. The filter can also accept changes specified by the operator to occur at specific points in the chromatogram.

The overall performance of this digital filter produced no measurable change in peak area measurements, and only depressed a Gaussian peak amplitude by 0.4%. It adjusts itself for variable-width peaks and simultaneously increases the peak detection sensitivity. The noise characteristics of the ADC are easily handled by the filter. The filter algorithm has an execution speed of about 20 msec per slope calculation in the microcomputer system, which allows enough time for execution of the real time peak detection program.

REFERENCES

- 1 R. Annino, *Advan. Chromatogr.*, in press.
- 2 G. M. Hieftje, *Anal. Chem.*, **44** (1972) 81A.
- 3 G. M. Hieftje, *Anal. Chem.*, **44** (1972) 69A.

- 4 A. Savitzky and M. J. E. Golay, *Anal. Chem.*, 36 (1964) 1627.
- 5 J. Steiner, Y. Termonia and J. Deltour, *Anal. Chem.*, 44 (1972) 1906.
- 6 S. N. Chesler and S. P. Cram, *Anal. Chem.*, 43 (1971) 1922.
- 7 S. N. Chesler and S. P. Cram, *Anal. Chem.*, 44 (1972) 2240.
- 8 M. Goedert and G. Guiochon, *Chromatographia*, 6 (1973) 76.
- 9 M. Goedert and G. Guiochon, *Chromatographia*, 6 (1973) 39.
- 10 H. C. Smit and H. L. Walg, *Chromatographia*, 8 (1975) 311.
- 11 A. C. Brown, D. L. Wallace and S. Kung, *Pitts. Conf. Anal. Chem. Appl. Spectr., Cleveland, Ohio, March 1-5, 1976*, Abstr. No. 336.
- 12 M. H. J. van Rijswick, *Philips Res. Rep., Suppl.*, No. 7 (1974) 32.
- 13 R. B. Blackman, *Linear Data-Smoothing and Prediction in Theory and Practice*, Addison-Wesley, Reading, Mass., 1965.
- 14 S. D. Stearns, *Digital Signal Analysis*, Hayden, Rochelle Park, N.J., 1975.