Sampling Frequency, Response Times and Embedded Signal Filtration in Fast, High Efficiency Liquid Chromatography: A Tutorial

M. Farooq Wahab[‡], Purnendu K. Dasgupta, Akinde F. Kadjo, Daniel W. Armstrong^{*}

Department of Chemistry and Biochemistry, University of Texas at Arlington, 700 Planetarium Place, Arlington, Texas 76019-0065, United States

Corresponding authors: ^{*}sec4dwa@uta.edu; [‡]mwahab@ualberta.ca

Abstract

With increasingly efficient columns, elute peaks are increasingly narrower. To take full advantage of this, choice of the detector response time and the data acquisition rate a.k.a. detector sampling frequency, have become increasingly important. In this work, we revisit the concept of data sampling from the theorem variously attributed to Whittaker, Nyquist, Kotelnikov, and Shannon. Focusing on time scales relevant to the current practice of high performance liquid chromatography (HPLC) and optical absorbance detection (the most commonly used method), even for very narrow simulated peaks Fourier transformation shows that theoretical minimum sampling frequency is still relatively low (< 10 Hz). However, this consideration alone may not be adequate for real chromatograms when an appreciable amount of noise is present. Further, depending on the instrument, the manufacturer's choice of a particular data bunching/ integration/response time condition may be integrally coupled to the sampling frequency. In any case, the exact nature of signal filtration often occurs in a manner neither transparent to nor controllable by the user. Using fast chromatography on a state-of-the-art column (38,000 plates), we evaluate the responses produced by different present generation instruments, each with their unique black box digital filters. We show that the common wisdom of sampling 20 points per peak can be inadequate for high efficiency columns and that the sampling frequency and response choices do affect the peak shape. If the sampling frequency is too low or response time is too large, the observed peak shapes will not remain as narrow as they really are - this is especially true for higher frequency for high efficiency and high speed separations. It is shown that both sampling frequency and digital filtering affect the retention time, noise amplitude, peak shape and width in a complex fashion. We show how a square-wave driven light emitting diode source can reveal the nature of the embedded filter. We discuss time

uncertainties related to the choice of sampling frequency. Finally, we suggest steps to obtain optimum results from a given system.

Keywords: High-efficiency chromatography, Fast separations, Sampling theorem, Response times, Digital filters

1. Introduction

The goal in analytical chromatography is to obtain baseline separations with maximum peak capacities per unit time without sacrificing detection sensitivities (and hence detection limits) and reproducibility. While the last parameter is largely governed by the fluid handling system, all the foregoing benefit from better chromatographic efficiencies. Achieving high efficiencies may be dependent on the column technology but actually visualizing this is also demanding of the detection and the data acquisition system. Efficient columns generally require smaller sample volumes. Ever smaller analyte amounts passing through the detector at ever faster flow rates creates a challenge for the detection system as well.

In the context of HPLC, the terms "fast", "ultrafast", etc. are yet undefined. However, a survey of the current literature suggests that both these terms imply sub-minute separations [1-5]. The efficiencies attainable by present separation media can only be realized if the rest of the system does not pose limitations. History suggests that the development of the rest of the instrumentation often lags behind advancements in column technology [2-4, 6, 7]. As early as the 1970s, Halász showed that while a baseline resolution of two components on a well-packed 5 μm silica column could be achieved in under 2 minutes, the existing HPLC hardware had to be modified to reduce extra-column effects [6].

In the last two decades, rapid developments have taken place in liquid chromatography. Smaller particles have enabled faster analyses by allowing shorter columns with the same efficiency as previous 25 cm benchmarks. The early 2000's saw a rapid increase in separation speed. For example, 7 analytes were separated in 24 s in reversed phase liquid chromatography [8], whereas

8 anions were separated under 30 seconds on monoliths in the ion-pairing mode [7]. The performance bottleneck was often the detection technology. With present well-packed core-shell silica columns, reduced plate heights (h) ≤ 2 are becoming common, even in columns of 5 cm length [3, 9]. Such separations are characterized by very narrow peak widths, very fast elution times and relatively low back-pressures. The core-shell morphology is often the preferred choice for HPLC/UHPLC instruments since it allows column efficiencies comparable to sub-2 μm fully porous particles while permitting the use of conventional HPLCs and column hardware. Recently, enantiomeric separations were reported on 3-5 cm long columns packed with 2.7 μm core-shell bonded chiral selectors in 4-15 s ($h = 1.6\text{-}2.0$) on a modified state-of-the-art UHPLC [3].

With high-speed separations and sub-2 h values, detector/data system parameters can significantly affect the observed results. Even today's state of the art high performance instruments can potentially limit the "true" performance of a chromatography column. In our experience as well as that of others [3, 5, 10-13], even current generation instrumentation can contribute as much as 30-60% to peak variance for short, narrow bore columns. Remedial focus is often on injection ports and connections tubing as these are easily changed. The influence of the choice of detector/data system parameters such as response time/time constant and data sampling/acquisition rate is often ignored. Few chromatographic publications discuss these and the chosen parameters are often not even stated despite the well-recognized importance of sampling frequencies and time constants in signal processing over the years. However, there is no consensus on mathematical approaches and/or goals. Table 1 summarizes the various suggestions/opinions of different individual investigators for selecting the proper sampling

frequency for chromatographic separations. It should be noted that many of the suggestions differ because the focus of the investigations differed. For example the goal some workers was to accurately render peak heights [14] or while others considered peaks areas [15, 16] and some focussed on the minimum number of points needed to describe a Gaussian function [17, 18]. Heien et al. suggested oversampling of data during signal acquisition up to MHz frequency followed by signal averaging [19]. Several texts suggest 20 points per peak [20]. Data collection rate having no influence on band broadening and resolution has been suggested as well [21] (Table 1). Herein we focus on the nuances associated with detector settings, namely, the data sampling frequency, the so-called detector response time, and the effect of embedded filters using three illustrative current generation HPLC instruments. Real and simulated examples are shown and recommendations are made.

2. Theory

2.1. Sampling frequency

The overall observed variance of a chromatographic peak σ^2 , calculated by the second statistical moment, is composed of the following contributions:

$$\sigma_{total}^2 = \sigma_{column}^2 + \sigma_{injector}^2 + \sigma_{tubing}^2 + \sigma_{detector\ volume}^2 + \sigma_{detector\ electronics}^2 \quad (1)$$

Equation (1) assumes that all the variances are independent of each other. Typically, the user has some control over the first three components. Less commonly, one can occasionally choose a particular detection cell but not the detector electronics and or any embedded signal processing algorithm. The user is generally not even aware of the latter, even though the consequences of this can be profound. For example, detectors based on aerosol generation (light scattering, charge detection, etc.) used in HPLC have a limited linear dynamic range (LDR), the signal being less

than linear at higher analyte concentrations. The firmware commonly applies a power function to the signal S and the output is S^y where $1 < y < 2$. This results effectively in a greater LDR and has the (un)intended consequence of making the chromatographic peaks to appear more efficient than they really are [22-24]. For the more commonly used optical absorbance detectors, the available choices in detector electronics typically allow a user to choose a particular data sampling rate (a.k.a. sampling frequency in Hz) and a response time or a time constant. The variance contributions from detector electronics can have a large impact on the peak (height, width and asymmetry) and retention time (see below).

All signal generation and acquisition is subject to the sampling theorem, which is known by various names. Whittaker first mathematically addressed the sampling theorem in 1915 for frequency band limited functions [25]. In 1926, Nyquist pointed out the importance of data sampling frequency but did not elucidate theory of the minimum sampling frequency. In 1933, Kotelnikov formally applied the theorem to communication engineering and Shannon investigated the sampling problem in detail in 1948. For an interesting history and citations to the original literature, the reader is referred to Luke [26]. Shannon [27] states in “*Communication in the Presence of Noise*”: “If a function $y(t)$ contains no frequencies higher than f cycles per second (Hz) it is completely determined by giving its ordinates at a series of points spaced $\frac{1}{2f}$ seconds apart.” This theorem is exact and not an approximation. This implies that if a signal is collected for a period t , $2ft$ points will completely describe the signal.

In the chromatographic context, the choice of a sampling frequency by an analyst then requires knowledge of the highest frequency component in a chromatogram that carries information (i.e.,

as opposed to noise), this will generally be associated with the narrowest peak in the chromatogram. As a first approximation, with present-day columns, most eluite peaks can be represented by a Gaussian function:

$$y(t) = \frac{1}{\sigma_t \sqrt{2\pi}} e^{-\left(\frac{t}{\sqrt{2}\sigma_t}\right)^2} \quad (2)$$

Where t is the time and σ_t is the standard deviation in time, the peak half width being given by $2\sqrt{2 \ln 2} \sigma_t$ or $2.355\sigma_t$ and a peak efficiency of $(t_r/\sigma_t)^2$, where t_r is the retention time of the peak. The frequency components (f) in the Gaussian peak are given by the Fourier transform of the Gaussian, $H(f)$ [28]:

$$H(f) = e^{-\left(\frac{t}{\sqrt{2}\sigma_f}\right)^2} \quad (3)$$

where the standard deviation of the Fourier transform in terms of frequency, σ_f is given by:

$$\sigma_f = 1/(2\pi\sigma_t) \quad (4)$$

The inverse relationship between standard deviation in the frequency domain σ_f and standard deviation in the time domain σ_t are both fundamental properties of Gaussian functions (see also Figure 1A and 1B) [29]. *Undersampling* and *oversampling* then refers to sampling below or above the sampling frequency ($2f$) as dictated by the Shannon theorem.

In practice, it should be noted that no peak is a perfect Gaussian or even an exponentially modified Gaussian. However, the Gaussian approximation is one of the best and most popular models for peaks on modern high efficiency chromatographic systems. Even extremely high efficiency columns with narrow peaks ($h < 2$), may exhibit some tailing [30]. This tailing arises

not only from the wall effects of the packed bed but also from the fundamental properties of laminar flow, which occurs in connecting tubing and the flow cell.

2.2. Noise filtering by time constants and response times

The above discussion does not account for noise present in any real signal. All present instruments seek to reduce the random noise, most commonly by digital filtering. Typically, the user can choose the time constant/response time from a list of available values; with some manufacturers, the choice of a response time results in automatic selection of a sampling frequency. When filtering is implemented in instrument firmware, it is relatively inflexible. When filtering is software implemented, it typically involves post processing and one has a choice of one or more filters, some common ones being Savitzky-Golay, moving average, Gaussian etc. (such post-processing data are not available in all chromatography softwares). The specific digital filter implemented in firmware is rarely specified by a manufacturer. Volumes have been written on various types of filters and their designs [31, 32]. Without knowing the nature of the specific filter present in the firmware, it is impossible to understand how a given system will respond to a step function. The term *time constant* (τ) applies where an asymptotic approach to the final state is involved, as with a resistor-capacitor (RC) based filter or its equivalent; this “*e-folding time*” (the time it takes to reach 63.2% of the step function, $(1 - \frac{1}{e}) = 0.632$, see Excel file in Supplementary Information). “Time constant” has little or no meaning with other types of filtering. One major HPLC manufacturer for example uses a Hamming window filter to reduce noise [33], whereas another uses a moving average filter (a.k.a boxcar filter/integrator) that includes a Gaussian weighting function [34]. It is rarely specified as to what a given manufacturer means by “response time” (a.k.a. the rise time), albeit the common

interpretation is that it is the time required for the signal to rise from 10% to 90% in response to an upward step function. This is the American Society for Testing and Materials (ASTM) definition [35] and we will use this definition of “response time” hereinafter. The exact output response will depend on the exact filter used, even if the 10-90% rise times are the same. Note also that some filters may be asymmetric; the 90-10% fall time observed with a downward step function may not be the same as the 10-90% rise time. For a RC type filter, the 10-90% rise time or 90-10% fall time equals 2.2τ . Aside from the nature of the filter, there is the choice of sampling frequency. If a signal is severely undersampled, a faithful rendition is obviously impossible; gross oversampling with no changes in filter conditions on the other hand will lead to increased noise (see below). While we assume that a given detector hardware/firmware and the data system is faithfully recording the eluate profile exiting the column, it is actually so only if the sampling frequency and filter conditions are appropriate. As we illustrate below, the present practice of “fast” HPLC can certainly test the capabilities of the rest of the system.

As to noise, in any detection system, the fundamental root mean square (rms) noise limit is the shot noise given by the Schottky equation [36]

$$\text{Shot noise (rms)} = \sqrt{2Ie\Delta f} \quad (5)$$

where I is the photocurrent, e is the elementary charge on the electron, and Δf is the bandwidth.

The response time is inversely proportional to the bandwidth.

$$t_{\text{response}} = \frac{\text{constant}}{\Delta f} \quad (6)$$

The constant is typically between 0.3 - 0.5 and depends on the characteristics of the filter used.

In chromatographic data acquisition, the lower bound of f is zero (DC) and the upper bound is dictated by the sampling frequency. If filtering does not bring in new noise sources, the rms noise increases with the square root of the sampling frequency. The Supplementary Information

provides simulation and brief explanation of three digital filters used by three major manufacturers, namely, numerically simulated RC filter, boxcar averaging and Hamming window filter along with Excel sheets.

2.3. Association of response or rise times with digital filters

As alluded in Section 2.2 time constants are meaningful only for filters with exponential decay type transfer functions *e.g.* low pass RC filters. Besides classical RC filters, other filters are also used in modern HPLCs which are finite impulse response filters (moving average, weighted moving averages with a Gaussian kernel etc.). So how do other digital filters in modern chromatography systems associate a response time with them? We illustrate this point with a simple moving average on a step function. The response time effect of filters is often studied with step functions. The most common step function is the Heaviside step function, which simply changes its value from 0 to 1 (see Figure 2) instantaneously. The general form of the moving average can be described by equation (7), where y_j^* is the smoothed ordinate value formed by averaging N integers. The index j is the running ordinate. The coefficient C_i is equal to unity in a simple moving average. In the case of weighted moving averages, the coefficient C_i is not equal to unity, but it can represent coefficients of a function (*e.g.* Hamming window uses a cosine function and is used by a major manufacturer of UPLCs). For a centered moving average, the value of m is odd.

$$y_j^* = \frac{\sum_{i=-m}^{i=m} C_i y_{i+j}}{N} \quad (7)$$

For example, a 5 point central moving average would be *e.g.* y_3^* = (average value of y_1 , y_2 , y_3 , y_4 , and y_5) and the “smoothed” data will be plotted at sampled time t_3 from corresponding times t_1 ,

t_2 , t_3 , t_4 , and t_5 . The moving average window is moved onward and the same process repeated. By a multi-pass moving average, we imply convoluting the original data more than once. For instance, output can be subjected to a 5-point moving average. The resulting “smoothed” data can be subjected to another pass of 5-point moving average. In Figure 2, we simulate a step function, which is being “smoothed” by a centered moving average. As the number of moving average passes increase, the step function becomes slower and slower i.e. it takes longer and longer to go from 10% to 90%. The readers can now associate a “response time” with moving average filter using the definition of response time: it is the time it takes for the output to rise from 0.1 to 0.9 or 10% to 90% of the original step function. Thus, the response time is an internally set parameter by the manufacturer with respect to a unit step function rather than a chromatogram. Instead of a multi-pass moving average, the digital filter can be any other type. The response time in each case, is the time it takes for the signal to rise from 10% to 90%. Illustrative examples of various filters on simulated chromatograms are shown in the Supplemental information.

3. Materials and Methods

3.1. Simulations and Calculation Software

Simulations of chromatograms, RC low-pass and moving average filtering were performed in Microsoft ExcelTM 2013; the Solver Add-in was often utilized [37, 38]. Gaussian white noise [39] was added to the simulated chromatograms by using the “Random Number Generation” function in Excel in Data Analysis option. Fast Fourier Transform (FFT) of Gaussian peaks and noise analysis was performed in OriginPro 2015 (Origin Lab Corporation, MA). For the Fourier analysis of the signal, the sampling interval and the raw response as a function of time seconds

was used for FFT (Signal Analysis feature in OriginPro). The time data (simulated or real) were assumed to be in seconds. The FT output is depicted in the magnitude vs. frequency (Hz) format.

3.2. Chemicals

Uracil, phenol, and acetonitrile were purchased from Sigma-Aldrich (St. Louis, MO). Deionized water (18.2 MΩ) was produced by a Millipore Synergy 185 system (EMD, Billerica, MS, USA).

3.3. Instrumentation and Chromatographic Conditions

Brand names are not explicitly specified here, as it is not our purpose to exalt or impugn any particular instrument; All instruments were controlled and data acquired with the current generation of the manufacturer's software. We used three HPLC systems, all are current models (see Supplementary Information). System 1 was a UHPLC system (manufacturer A, 120 MPa max), System 2 was a HPLC system (manufacturer A, 60 MPa max). The UHPLC was used for chromatography and the HPLC was used for square wave experiments only. The UHPLC has been optimized for plumbing and minimum extra-column volume possible in our laboratory. Namely, the injector-column and column detector connections were made with minimum possible lengths of 75µm i.d. polyether ether ketone (PEEK)-coated fused silica tubing with nanoViperTM fittings (www.thermoscientific.com). This may not be the standard format in typical laboratories. Secondly, the standard injector has been replaced by a low volume injector and needle seat.

System 3 was a HPLC system (manufacturer B, 66 MPa max). All three instruments were equipped with diode array absorbance detectors recommended by the manufacturer for the particular instrument. The third system was used as configured by the manufacturer without

reducing extra-column tubings in the auto-sampler and detector. The auto-sampler has a 50 μ L injection capability. In System 3, the injector-column and column-detector connections were made with minimum possible lengths of 75 μ m i.d. nanoViper fittings as detailed above.

All chromatographic experiments were conducted with a Poroshell-120 Column (120 EC-C18, 2.7 μ m, 150x4.6 mm i.d., P/N 693975-902, www.agilent.com). According to the manufacturer's certification, this particular column exhibited 38000 plates for their test chromatogram. We used isocratic elution with 80% acetonitrile and 20% water as eluent at a flow rate of at 1.8 mL/min ($\Delta P = \sim 36$ MPa) for a separation of uracil and phenol at ambient temperature on system 1 and 3 (see Supplementary Information for more details). The injection volume in each case was 1 μ L and the detection wavelength was 254 nm.

3.4. Detector Response Time/Time Constant Experiments

Manufacturer A and B respectively refer to the choice of the response speed in their software as “response time” and “time constant”. To test the effects of the choice of these parameters, we affixed a small surface mount red-orange (λ_{peak} 610 nm) light emitting diode (LED) on the light source side of the flow cell with tape and the cell was then reinstalled in place. Lead wires soldered to the LED were connected to a function generator (5MHz, B&K Precision model 4011A) and driven by a square wave at 1 Hz.

4. Results and Discussion

For digital signals, the lower limit for sampling the signal is set by the sampling theorem, but there is no upper limit for the sampling frequency except as set by the capability of the analog to

digital converters (ADC's) [19]. Such conversion capabilities presently extend to the several GHz range [40], clearly far beyond the needs of HPLC detection. We focus here mainly on the effect of detector settings chosen by the user (sampling frequency and response times) on present-day fast and narrow chromatographic peaks.

4.1. Revisiting the Sampling Frequency Concept

In Figure 1 (A) we simulate three different chromatographic peaks of equal area but differing in peak widths ($\sigma_t = 0.08, 0.2$ and 1.5 s). A Fourier transform can identify what frequency components are present in such signals; this is also depicted in Figure 1 (B). For the narrowest peak ($\sigma_t = 0.08$ s, half peak width $w_{0.5} = 120$ ms), the FT is wide and conversely, the widest peak ($\sigma_t = 1.5$ s) has the narrowest frequency distribution, as predicted by the relation $\sigma_f = 1/(2\pi\sigma_t)$. Equation 4 thus provides the fundamental perspective why narrow peaks must be sampled at higher data collection rates.

Note that the Fourier transform of a Gaussian peak is also a Gaussian function [29]; the normally distributed amplitude in the frequency domain has its maximum amplitude at $f = 0$ (DC) and extends to $f = \infty$. It is readily estimated from standard distribution tables that 95%, 99%, and 99.9% of the information is then contained within $f = 1.645 \sigma_f$, $2.33 \sigma_f$, and $3.08 \sigma_f$, or in terms of the half peak width, $0.62/w_{0.5}$, $0.87/w_{0.5}$, and $1.15/w_{0.5}$ (using the relation $w_{0.5} = 2.355 \sigma_f$). An adequate sampling frequency ($2f$) for the different situations will range from 0.35 Hz for the widest peak while retaining 95% of the information content to 20 Hz for the narrowest peak with 99.9% information retention. The default sampling frequency seen in many instruments is 2.5

Hz, which is at least double the sampling frequency minimally needed to retain $\geq 95\%$ of the information on peaks with half-width ≥ 1 s.

The above discussion clearly suggests a different conclusion compared to a recent publication that posits sampling frequency should not affect the signal and even when *undersampling*, the original chromatogram can be readily recovered [21]. This concept will be revisited later in Section 4.2. This aside, the more important question is whether sampling at the minimally acceptable frequency according to the sampling theorem is adequate in the presence of noise.

Noise is to be found in any real measurement and typically, the noise has components at higher frequencies than is present in the analytical signal. Data systems of course do not have any innate intelligence to distinguish noise from the signal, even when a particular incidence of spurious noise would be readily apparent to the user upon examination of the data. Consider a chromatographic peak recording which has had an incidence of a spike during the chromatogram due to whatever electromechanical reason or an errant bubble. Consider also that the spike occurred shortly before the peak apex and its magnitude was enough to put it above the peak apex. If the sampling frequency is high enough, the spike will readily appear for what it is, whereas in the case of an undersampled or just adequately sampled system, if the spike coincided with a sampled interval, this point would be considered the peak apex and both the perceived peak height and the retention time will be inaccurate. While it is known that some filters can circumvent this problem (the Olympic filter [41] for example removes the highest and lowest value within a chosen window), it is generally easier to distinguish the signal from noise as the sampling frequency is increased.

Now, let us consider a simulated chromatogram (Figure 3) with four high efficiency peaks ($N = 13,000$ - $15,000$) with a total retention time < 1 min, realistically attainable on present-day short well-packed core-shell or sub- $2\ \mu\text{m}$ particle containing columns. The first peak at ~ 16 s represents a relatively weak signal, the second and third peaks are partially resolved, followed by a fourth peak at ~ 29 s. White Gaussian noise (mean zero, standard deviation 0.06 amplitude units) was added to the chromatogram. Panels A-D respectively depict the chromatograms as observed by sampling at 160, 40, 4, and 2.5 Hz, respectively, sampling frequencies available in most present day HPLCs/UHPLCs. Note that the minimum sampling frequency for the widest ($w_{0.5} = 0.61$ s) to the narrowest peak ($w_{0.5} = 0.31$ s) would theoretically be 2-4 Hz for 95% accurate reproduction. Figures 3A and 3B thus represent oversampled chromatograms, 3C is just adequate according to the sampling theorem and 3D represents a sampling frequency that is nominally adequate for the widest peak but not for the narrowest. It is easy to see that in the last case, much of the information is lost or distorted, the first peak is nearly gone, resolution between peaks 2 and 3 is worse, and none of the peak apices is correctly registered. Undersampling compromises the peak height and areas. In contrast, by oversampling (160 Hz), the original signal is collected with high fidelity; no information is lost during the sampling process. However, Figure 3A makes it apparent that this is obtained at a cost of increased baseline noise.

The noise in the raw data is the same on the input signal in all four chromatograms. The inset in Figure 3 shows the local noise from 0 to 10 s. So what causes higher noise at the highest sampling frequency? This results from purely statistical reasons and the sampling theorem also predicts the same. The noise amplitude distribution is Gaussian, by increasing the sampling

frequency; we increase the probability of picking the highest amplitude values. Statistics also suggests that had we considered a proportionately longer interval to consider the noise (such that the number of points in the evaluated window remains constant as a function of frequency), the noise would have been the same. Thus, an alternative perspective is that by lowering the sampling frequency, we exclude noise present at higher frequencies. However, the concept of *local noise* is important (this has been so invoked already in image processing [42]) because elution of a peak is a local event. The narrower the peak, the more localized it is in the time domain and more it is affected by *local noise*. The effect can be especially deleterious at low signal amplitudes and can clearly be seen for the first peak in the 160 Hz chromatogram. In an oversampled chromatogram, *assuming all electronic effects and data massaging are absent*, the lowest amplitude peaks are more likely to be affected by noise. At a sampling frequency of 40 Hz, the local noise level has decreased sufficiently that the earliest eluting peak is demarcated with less noise. At 4 Hz, the last peak height is lower compared to those obtained at higher sampling frequencies and at 2.5 Hz it is lower still. Figure 4 shows a magnified view of the first peak at different sampling frequencies. It is important to note that simulating sampling frequencies as represented here by picking every n th point from the raw data may not represent how the noise is affected by the sampling frequency choice in present data systems. This is a function of the nature of the ADC used (for a summary of principal types of ADCs see [43]). A Delta-Sigma ADC, the most common type used in data acquisition systems, intrinsically uses a fixed sampling frequency, the highest it is capable of. Rather than picking every n th point, an n -fold lower requested rate results in integration/bunching of n data points with consequent lowering of the noise. Figure 4 also shows what the 40 Hz chromatogram is more likely to

appear, where in a master 160 Hz data, every successive 4 point bunch has been averaged, resulting in a ~2-fold decreased noise compared to the original representation.

Aside from noise, the sampling frequency also affects the computed peak efficiency by missing the peak apex. For the 4th peak, for example at 29.4 s, there is loss of 7.7% in the peak height (peak apex missed by sampling at 4 Hz) for Fig. 3B vs. 3C. The peak efficiency, the peak shape, the exact retention time and their reproducibility are important parameters to a chromatographer and all benefit from some oversampling compared to that minimally required by the sampling theorem. In locating a peak apex, typically the highest point is simply chosen, there are no interpolations or fitting to some preconceived shape. The peak apex thus corresponds to a particular sampled point, whereas the real apex may be at a slightly different time. At a chosen sampling frequency of F Hz, the sampled points are $1/F$ s apart. It follows that the maximum error will be half the sampling interval or $1/2F$ s, purely due to the chosen sampling frequency. Thus, sampling at 160 Hz, the worst-case scenario will give us a retention time uncertainty of ± 3.1 ms whereas sampling at 2.5 Hz, the uncertainty is ± 200 ms.

4.2. Real chromatograms

In the following sections with real chromatograms, the term efficiency will refer to *observed* efficiency on the instrument due to detector settings. In Figure 5, we show four sub-minute high efficiency separations of uracil (peak 1) and phenol (peak 2). The chromatograms were obtained at a fixed $\tau = 0.01$ s but at sampling frequencies ranging from 0.5 to 100 Hz. The highest efficiency is seen at the highest sampling frequency but the noise is also the highest. There is a consistent decrease in efficiency and noise with decreasing sampling frequency as depicted in the Figure. At 40 Hz, one observes perhaps the best compromise between efficiency and noise is

realized although this is admittedly a subjective choice. Note that calculation of the peak efficiency has an uncertainty that also increases with decreasing sampling frequency, see the Supplementary Information for illustrative calculations. From >20000 plates at 100 Hz, if we sample at 6.25 Hz, which is approximately twice the minimum sampling frequency putatively needed for a peak of 0.92 s half-width (sampling theorem requires ~ 4 Hz as the bare minimum), the efficiency decreases to $\sim 14,000$ plates (Figure 5C). Sampling at even lower frequencies (Figure 5D) results in irreproducible chromatograms where both peak height and shape change from run to run.

A detailed analysis of noise will indicate that the decrease in noise in real systems is not always exactly predictable. Referring to the 10-30 s period, the baseline standard deviation was 43, 13, 13, and 13 units at 100, 40, 6.25, and 0.5 Hz. The noise decreases by a factor significantly more than the expected amount in going from 100 to 40 Hz. This suggests that additional sources of noise appear at the higher frequencies (i.e., unlike our illustrative example it is not white Gaussian noise) and below 40 Hz, the noise actually decreases very little, if at all, suggesting there are non-random sources of noise (and/or drift; computation of noise as standard deviation counts drift as noise). Remember however, τ was not nominally altered in the above experiments although when τ is smaller than the sampling interval (as in the case of the two lower sampling frequencies), the precise choice of time constant has no real effect.

Given the drastic decrease that we observe above in the peak efficiencies, one may well wonder why a recent publication stated: “When the data acquisition frequency is too low, only a few points per peak are recorded. We will miss the peak apices,..., but *peaks remain as sharp as they*

were at the highest sampling frequency” [21]. The reason is that these authors used a column with 2000-3000 plates only, and in such a case, it may not matter. The decision of the sampling frequency clearly will be dependent on the column performance level. A micrometer is not essential in measuring the height of a giraffe. In the same vein, gross oversampling, while certainly technologically feasible, has few advantages, aside from increased noise, results in very large data files, and increased noise level.

4.3. Effect of Noise Filters on the Signal

Noise is present with any real signal. As discussed, the choice of a high sampling frequency with an adequately fast response time provide the most efficient peaks, but with greater noise.

Whenever the signal bandwidth is known to be smaller than the noise bandwidth, noise at frequencies greater than those occurring in the signal can be attenuated, by appropriate processing, also called *filtering a signal*. Butterworth famously stated (some 30 years before such filters were even attempted): “An ideal electrical filter should not only completely reject the unwanted frequencies but should also have uniform sensitivity for the wanted frequencies” [44].

The simplest traditional filter used in analytical instrumentation is the RC filter where the output signal is connected to ground via a serial combination resistor (R) and followed by a capacitor (C). The output signal is collected from the junction of the R and C elements (see Supplementary Information). When R is in Ohms and C is in Farads, $\tau = RC$ and is in seconds. It can be readily changed by changing the values of one or both components. An RC filter does have an essentially flat response in the desired frequency domain but higher frequencies are not abruptly cut off. Rather, they are exponentially attenuated as the frequency increases. Exactly how the output to input voltage ratio changes as a function of frequency is referred to as the *transfer*

function of the filter. An RC filter is a “single-pole low-pass filter”. For a short summary on filters and “poles” see reference [45]. A “pole” is so-called because if the transfer function is visualized in the Laplace domain, imagining the response surface as a flat rubber sheet, it will stick up in a particular frequency region as if supported by a pole underneath. Much sharper cut-off at higher frequencies than the exponential *roll off* provided by a simple RC filter are possible and may be desirable in many cases. Whereas hardwired electrical circuits with a multitude of elements were used in yesteryears, currently signal filtration is invariably carried out after digitization. The digitized data is typically processed by the desired mathematical algorithm implemented in instrument firmware; hence, they are called digital filters. Some digital filters may have a transfer function identical to a RC filter. In other cases, the simplest of digital filters, a “moving average” filter, is very easy to implement in the digital domain but not on a hardwired basis. In a simple n -point moving average filter (a function built into data analysis packages of common spreadsheets like Microsoft ExcelTM), the input is n successive points in the data array and the output is their average – the algorithm then simply moves down one row in the data array and repeats the process. The simplest of moving average filters is said to have a *rectangular kernel*, the rectangle represents the filter window while the flat top of the rectangle indicates that all the points in the chosen window are weighed equally in the averaging process. In other variants of the moving average filter, individual elements within the averaging window may be weighed differently, e.g., by a Gaussian function centered at the center of the averaging window. Running the signal through the same moving average filter multiple times is also possible. Running it twice through the same rectangular moving average filter is essentially equivalent to a triangular filter kernel whereas 4 or more passes of a rectangular kernel results essentially in a Gaussian filter kernel (See Figure 2). An excellent discussion of moving average filters appears

in reference [46], which also categorically states “*In spite of its simplicity, the moving average filter is optimal for a common task: reducing random noise while retaining a sharp step response. This makes it the premier filter for time domain encoded signals. However, the moving average is the worst filter for frequency domain encoded signals, with little ability to separate one band of frequencies from another.*”

For a simple moving average filter, if a data points are averaged in the window and the sampling frequency is $2f$ Hz it is easily verified that the response time is $0.4a/f$ s. It is interesting to note that running a moving average filter n successive times increases the response time by a factor slightly less than $\sqrt{n}$. A moving average filter can be implemented in a nested configuration.

While the simple moving average filter is extremely fast in execution, modified moving average filters can be orders of magnitude slower in computational execution albeit they do a much better job in moving frequency domain noise [40]. Fortunately on an absolute scale even the highest sampling frequencies in current HPLC detectors is relatively low and implementing almost any filtering algorithm does not pose a problem with present microprocessors.

The effect of signal filtration on chromatographic or spectroscopic signals has been previously examined [47, 48]. However, most of this literature in the analytical chemistry context is dated and sometimes contradictory [21, 48] (Table 1). Filters of yesteryears were invariably RC types and have been discussed (*vide supra*). Hinshaw discussed for example artifact tailing that is observed if an inappropriately high τ is used [48].

Presently, we examine the effect of the chosen response time (manufacturer A) or time constant (manufacturer B) on our illustrative systems by using a LED source driven by a 1 Hz square

wave, to simulate a sharp peak of 0.5 s “width”, comparable to that of the narrowest peaks attainable from present high efficiency columns. Figure 6 shows the results; note that there was no additional superimposed noise and all sampling frequencies were well above the lower limit dictated by the sampling theorem to capture the repetition frequency (but not the sharp rising or falling edge of a square wave, which theoretically extends to infinity but because of the function generator limitations and LED capacitance is likely to be several hundred kHz. Manufacturer A (system 2, panels A-D) picks the sampling frequency as soon as the response time is chosen, they cannot be independently chosen. At 80 Hz sampling frequency, the square wave is reasonably reproduced with 31 and 63 ms response times although rabbit ears (overshoots) begin to appear in the latter case at the edges but not very reproducibly (Figures 6A and 6B). At 40 Hz and 130 ms response time (Fig. 6C), the peak shape is distorted, is irreproducible and occasionally has minor asymmetry but at still lower sampling frequencies and slower response times (Fig. 6D) the signal is grossly over filtered; it regains symmetry, assumes a more Gaussian character and essentially appears as a sine wave. While the manufacturer does not state the specific filter used, a moving average filter with an appropriate kernel (Gaussian, or Hamming), or multiple passes of a moving average filter can reproduce this behavior (see Section 2.3 and supplementary simulation).

The right panel of Figure 6 shows the behavior of system 3 for a similar experiment. In this case, the sampling frequency was maintained constant and τ was varied. With $\tau = 10$ ms, the square wave was faithfully reproduced (Fig. 6E). At higher τ (ranging from 160 to 2000 ms) the choice of the sampling frequency was immaterial as the behavior is governed by the choice of τ . At 160 ms (Fig. 6F) the response has become seemingly asymmetric with a peak that is tailing. In reality, however, the rise and fall behavior are mirror images (along y-axis). Again with

continued increase in τ , symmetry is regained and eventually a triangular wave of lower amplitude results (Figures 6G and 6H). This filter behavior is that of a classical RC filter and the manufacturer reference to “time constant” is thus appropriate.

4.4. Effect of filtering parameters on fast chromatography

We assessed the effect of filter parameters on System 3, which allows the sampling frequency and the response time to be independently chosen. Several major manufacturers, (Dionex, Shimadzu, and Waters, to name a few) offer a similar choice and leave the user to make the judgment that a particular combination is judicious. In Figure 5, we had fixed τ at a very low value and gradually decreased the sampling frequency to a point where the latter became the limiting factor and were clearly not a judicious choice. Presently we fix the sampling frequency at 100 Hz and vary τ over a wide range from 0.00 (no filter) to 2.00 s, using identical conditions to the experiment in Figures 6E - 6H. The results are shown in Figure 7. Predictably, baseline noise decreases monotonically with increasing τ but peak efficiency also decreases, as labeled in the Figure. The last two choices are again clearly not judicious. Peak asymmetry (in terms of tailing factor) also increases with increasing τ , respectively being 1.0₅, 1.1, 1.5 and 2.5₅ in Figures 7A - 7D for the second peak. Less apparent are the decreases in peak height (From Figure 7A to 7D the height has decreased by a factor of 66%) and the perceived retention time (peak apex) that shifts to larger values as the τ is increased. Note that time constants should not affect peak areas. The change in area in going from a time constant of 0.00 s (no filter option) to 2.00 s is only 3.5 %, is accountable entirely from injector reproducibility since the injection volume is only 1 μ L (kept low to minimize band spreading). In a RC type low-pass filter, the relative width of the distorted peak depends only on the ratio of the τ to the original width of the peak [47]. At the highest τ chosen, the plate count for phenol has decreased to 2700 from the

original 22200 (Figure 7 A to D) accompanied by severe apparent tailing, both total artifacts brought about by the data system parameters.

Instrument operation manuals often correctly suggest that a smaller τ be chosen if peaks are narrow, without specific recommendations as to what for example an appropriate ratio of the two should be. Such problems are compounded by popular tools that are supposed to simulate the behavior of an HPLC system. In the “HPLC Simulator” [49], an otherwise admirable tool, when the “time constant” is increased, the peaks simply broaden symmetrically. Obviously, that is not what will be observed if a RC type filter is present (which would be suggested by the use of the term time constant).

Meyer has suggested a method for choosing τ if retention time and efficiency are known beforehand, if θ is the allowed percent broadening of the peak width [50]:

$$\text{Time constant} \leq \theta \left(\frac{t_r}{\sqrt{N}} \right) \quad (7)$$

If 1% broadening of the peak width is acceptable, for the values for phenol in Fig. 7A ($t_r = 53$ s; $N = 22000$), with peak efficiencies of 22000 and retention time of 52 s (phenol), the calculated time constant would be 0.357 s. In reality, even with a time constant of 0.160 s, we observed that N decreases to 17700 plates, representing a >2% increase in peak width. For a Gaussian peak, it is possible to predict how $w_{0.5}$ will change as a function of $\tau/w_{0.5}$. Table 2 lists values the increase in $w_{0.5}$ and the peak asymmetry at 5% and 10% of peak height after processing through an RC filter as a function of $\tau/w_{0.5}$. Note that when we use the dimensionless parameter $\tau/w_{0.5}$ rather than the absolute value of τ , the effects on an initially ideal Gaussian peak becomes independent of the $w_{0.5}$. Alternative to using or interpolating values in Table 2, the Excel workbook RC filter

Gaussian.xlsm given in the Supplementary Information accepts $w_{0.5}$ for the raw peak (assumed to be Gaussian) as the input and provides the output for any time constant chosen, automatically generating all the peak parameters of interest listed in Table 2. Another version of the same program, *RC filter anydata.xlsm* allows inputting the actual x - y data array for a peak, which may not be Gaussian.

4.5. Effects on best case performance

With minimized extra-column connection volumes, System 1 (the UHPLC System) attained 39000 plates compared to manufacturer's certification of 38000 plates minimum of this particular column batch. As previously stated a choice of the response time in this system automatically results in a choice of the sampling frequency that ranges from 160 Hz to 0.31 Hz. A choice of 0.031 or 0.063 s as response time will automatically result in a sampling frequency selection of 160 Hz, a choice of 0.13 s, 0.5 s or 4.0 s response times will automatically result in sampling frequencies of 40, 10 and 1.25 Hz respectively. This approach prevents the user from making a choice of an inappropriate combination but does not of course prevent the choice of an altogether inappropriate response time that has an equally inappropriate associated sampling frequency. The coupling of appropriate sampling frequencies and response times is specifically mentioned in the manuals of the manufacturer. Calling this "undocumented feature of the software, which can deceive the users..." is not accurate [21]. The vendor emphasizes in its documentation that choosing a lower sampling frequency does not actually change the intrinsic sampling frequency but the output data are simply averaged/integrated in bunches (see previous discussion on Delta-Sigma ADCs).

Referring to Figure 8, based on the narrowest peak (Figure 8A, $w_{0.5}$ 0.6132 s; note that we also measured with a response time of 16.0 ms and a sampling frequency of 160 Hz, there was no discernible difference with Fig. 8A), all of the sampling frequencies used should be almost adequate, so the observed changes are due solely or dominantly from changing the response time. However, with such highly efficient peaks even the accuracy of determining retention time plays a role that is not often appreciated. To determine half-width, the temporal interval between two actually measured points will generally be taken and no interpolation used. In going from 80 Hz to 40 Hz, the uncertainty in locating each point goes from ± 6.25 ms to ± 12.5 ms. That for measuring an interval will be a factor of $\sqrt{2}$ higher: at 40 Hz the overall uncertainty is 17.7 ms. For the present parameters, ($t_r = 52$ s, $w_{0.5} = 0.6132$ s, an uncertainty of 17.7 ms results in uncertainty of ± 2200 plates in the value of N! In comparing Figures 8A and 8B (39000 vs 37940), this needs to be borne in mind that the small difference that exists can be due both to the increases in the sampling interval and the response time. In going to Fig. 8C, it is interesting to note that from between its 1% bounds, this peak comprises of 20 sampled points, a number often prescribed as adequate [20]. The sampling frequency may be adequate to describe the peak shape (sampling points 100 ms apart) but a response time of 500 ms is clearly inadequate for a $w_{0.5}$ only slightly higher than this and as a result, there is a near 50% loss in efficiency compared to the best case. The choice of sampling/response parameters in Figure 8D is simply a gross disservice to this highly efficient column. It essentially goes to show that a manufacturer may provide compatible combinations of response times and sampling frequencies but the determined analyst can always find a way to pick the most inappropriate combination!

There are both commonalities and differences between the results in Figures 7 and 8. In both cases, as τ or response time increases noise decreases in a consistent fashion. However, unlike in

Figure 7, the increase in t_r with increasing response time was minimal in Figure 8. Also unlike Figure 7, increases in the response times did not induce peak tailing. These results are consistent with the results of the previously described square wave experiment, and as stated therein, consistent with the use of a Gaussian kernel (or multi-pass centered moving average filter).

4.6. Suggestions for practitioners and manufacturers

In modern HPLC instrumentation, details of detection and signal processing are largely black boxes to most users. Details of signal processing are rarely, if ever, available to the analyst. The choice of data acquisition and signal processing parameters within a given instrument and the choice of different signal processing techniques between different instruments can affect the chromatographic output data, especially for present day fast separations. Those engaged in stationary phase development and column-packing processes need to be aware of not only extra-column effects but also of detector settings. These effects become significant with short narrow bore columns, analytical columns that are highly efficient ($h < 2$) and/or ultrafast separations. The following are our suggestions to the practicing chromatographers:

(i) Use the test mixture, injection volume and chromatographic conditions used by the manufacturer to test the column efficiency. Use the highest sampling frequency and lowest value of τ or response time that your system will allow. Use a sufficiently high analyte concentration so that visualization of peaks or efficiency calculations is not affected by the baseline noise. If you do not get close to the manufacturer specified efficiency, minimize all connection tubing diameter and length after the injector and make sure that connections themselves have no dead volume. This will be the best you can do. Run samples of your interest under appropriate chromatographic conditions that you have developed. If you are not bothered by the present

noise level or the data file size of your chromatograms for your samples of interest, you can leave the data acquisition settings at their fastest but most noisy level.

(ii) Consider the half-width of the narrowest peak in your chromatogram. If you can afford to sacrifice some efficiency, estimate what increase in $w_{0.5}$ will not affect your quantitation (remember that efficiency will change with the square of the change in $w_{0.5}$, 1% and 10% changes in $w_{0.5}$ respectively connote a 2% and 21% change in efficiency. Table 2 will guide you to the τ , you might use. If your system uses response times, estimate on the basis that response time equals 2.2τ . If your system chooses the sampling frequency based on the response time you have chosen then no further action is needed. If you need to pick a sampling frequency, you can proceed in one of two ways. If the maximum error in retention time accuracy you can tolerate is x seconds, pick a sampling frequency (Hz) of $1/2x$ or higher. Alternatively, if you know the number of points (n) across the (narrowest) peak you are comfortable with (while 15-20 is commonly suggested for n , we prefer *at least* 30 if high precision retention time or efficiency information is desired) pick a sampling frequency (Hz) that is no less than $0.5n/w_{0.5}$.

(iii) We suggest to the HPLC vendors to provide the user the ability to process different parts of the chromatogram with different filtering capabilities. It is common to have a “General Sensitivity Problem” analogous to the well-known “General Elution Problem” that while the chromatogram has some narrow tall peaks that are far above the noise levels it also has some broad low intensity peaks that are difficult to quantitate because of the noise. The post-processing the data can be readily done by exporting the data to a spreadsheet and applying different filters in selected regions of the chromatogram. The individual segments can be highlighted and different degrees of filtering (time constant/response time/bunching/variable

kernels or number of passes of a moving average filter) can be chosen for each segment. This will be particularly valuable especially if that can be then incorporated into a particular “method”. When that “method” is executed, this mode of data processing will be automatically implemented and it will perform quantitation based on calibration that has been carried out using the same “method”.

(iv) Researchers need to be aware that while all instruments offer the choice of sampling frequency and some filtering parameters, the results of comparable manipulations may vary from one instrument to another. As was demonstrated, increasing τ or response time may lead to tailing in one case and symmetric broadening in another. While the term time constant decidedly implies the use of an RC filter or exponential decay type transfer function, the use of the term “response time” can be applied to any filter type and the nature of the filter used cannot be determined from this specification. As we have seen in Figure 5, it is not possible to predict how noise will decrease with the sampling frequency in real instruments, as the nature of the noise in a real system is not properly described by white noise or for that matter, any pre-specified noise spectrum. Our conservative recommendation for high fidelity recording will be to use a *minimum* sampling frequency equal to $2/\tau$ or $4.4/(\text{response time})$

(v) To determine the time constant/ response characteristics of the detector, the ASTM suggests adding a UV absorbing component in the mobile phase and monitoring the response at various flow rates. Most detector configurations will allow the use of a simpler approach, the use of a small solid-state light source to provide a step function to the light received by the detector. This approach also reveals far more about the overall data processing method, which can reveal what is happening by changing the digital filtering times.

Acknowledgments

DWA thanks the Robert A. Welch Foundation (Y-0026) for support of this work. PKD acknowledges support from National Science Foundation through CHE-1506572 and from NASA through NNX11AO66G. We also thank the Advanced Analytical Instrumentation Center at the University of Texas at Arlington for the use of their facility.

References

- [1] S. Fekete, I. Kohler, S. Rudaz, D. Guilleme, Importance of instrumentation for fast liquid chromatography in pharmaceutical analysis, *Journal of Pharmaceutical and Biomedical Analysis*, 87 (2014) 105-119.
- [2] C.A. Lucy, M.F. Wahab, *Advances in High-Speed and High-Resolution Ion Chromatography*, LC GC Europe, (2013) 38-42.
- [3] D.C. Patel, Z.S. Breitbach, M.F. Wahab, C.L. Barhate, D.W. Armstrong, Gone in Seconds: Praxis, Performance, and Peculiarities of Ultrafast Chiral Liquid Chromatography with Superficially Porous Particles, *Analytical Chemistry*, 87 (18) (2015) 9137–9148.
- [4] J.C. Giddings, R.A. Keller, *Extracolumn contributions to chromatographic band broadening*, Marcel Dekker, New York, 1966.
- [5] C.L. Barhate, M.F. Wahab, Z.S. Breitbach, D.S. Bell, D.W. Armstrong, High efficiency, narrow particle size distribution, sub-2 μm based macrocyclic glycopeptide chiral stationary phases in HPLC and SFC, *Analytica Chimica Acta*, 898 (2015) 128-137.
- [6] H. Elgass, H. Engelhardt, I. Halász, Reproduzierbare Methode zum Packen von Trennsäulen für die Chromatographie mit Kieselgel (5–10 μm), *Fresenius' Zeitschrift für analytische Chemie*, 294 (1979) 97-106.
- [7] P. Hatsis, C.A. Lucy, Ultra-fast HPLC separation of common anions using a monolithic stationary phase, *Analyst*, 127 (2002) 451-454.
- [8] J.J. Kirkland, Ultrafast Reversed-Phase High-Performance Liquid Chromatographic Separations: An Overview, *Journal of Chromatographic Science*, 38 (2000) 535-544.
- [9] S. Fekete, K. Ganzler, J. Fekete, Efficiency of the new sub-2 μm core–shell (Kinetex™) column in practice, applied for small and large molecule separation, *Journal of Pharmaceutical and Biomedical Analysis*, 54 (2011) 482-490.
- [10] F. Gritti, S.J. Shiner, J.N. Fairchild, G. Guiochon, Evaluation of the kinetic performance of new prototype 2.1 mm narrow-bore columns packed with 1.6 μm superficially porous particles, *Journal of Chromatography A*, 1334 (2014) 30-43.
- [11] T.H. Walter, R.W. Andrews, Recent innovations in UHPLC columns and instrumentation, *TrAC Trends in Analytical Chemistry*, 63 (2014) 14-20.
- [12] K. Broeckhoven, G. Desmet, The future of UHPLC: Towards higher pressure and/or smaller particles?, *TrAC Trends in Analytical Chemistry*, 63 (2014) 65-75.
- [13] D.A. Spudeit, Z.S. Breitbach, M.D. Dolzan, G.A. Micke, D.W. Armstrong, Superficially Porous Particle Based Hydroxypropyl- β -cyclodextrin Stationary Phase for High-Efficiency Enantiomeric Separations, *Chirality*, 27 (2015) 788-794.

- [14] P.C. Kelly, G. Horlick, Practical Considerations for Digitizing Analog Signals, *Analytical Chemistry*, 45 (1973) 518-527.
- [15] N. Dyson, Peak distortion, data sampling errors and the integrator in the measurement of very narrow chromatographic peaks, *Journal of Chromatography A*, 842 (1999) 321-340.
- [16] S.N. Chesler, S.P. Cram, Effect of Peak Sensing and Random Noise on Precision and Accuracy of Statistical Moment Analyses from Digital Chromatographic Data, *Analytical Chemistry*, 43 (1971) 1922-1933.
- [17] B.G. Vandeginste, D.L. Massart, L.M. Buydens, S.D. Jong, P. Lewi, J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics: Part B. Volume 20 B in Data Handling in Science and Technology* Elsevier, Amsterdam, 1998.
- [18] P.J.H. Scheeren, Z. Klous, H.C. Smit, D.A. Doornbos, A software package for the orthogonal polynomial approximation of analytical signals, including a simulation program for chromatograms and spectra, *Analytica Chimica Acta*, 171 (1985) 45-60.
- [19] N.D. Laude, C.W. Atcherley, M.L. Heien, Rethinking Data Collection and Signal Processing. 1. Real-Time Oversampling Filter for Chemical Measurements, *Analytical Chemistry*, 84 (2012) 8422-8426.
- [20] Stavros Kromidas, H.J. Kuss, *Quantification in LC and GC: A Practical Guide to Good Chromatographic Data*, Wiley VCH Verlag, Weinheim, 2009.
- [21] A. Felinger, A. Kilar, B. Boros, The myth of data acquisition rate, *Analytica Chimica Acta*, 854 (2015) 178-182.
- [22] P.K. Dasgupta, Y.J. Chen, C.A. Serrano, G. Guiochon, H.H. Liu, J.N. Fairchild, R.A. Shalliker, Black Box Linearization for Greater Linear Dynamic Range: The Effect of Power Transforms on the Representation of Data, *Analytical Chemistry*, 82 (2010) 10143-10150.
- [23] R.A. Shalliker, P.G. Stevenson, D. Shock, M. Mnatsakanyan, P.K. Dasgupta, G. Guiochon, Application of power functions to chromatographic data for the enhancement of signal to noise ratios and separation resolution, *Journal of Chromatography A*, 1217 (2010) 5693-5699.
- [24] D. Shock, G.R. Dennis, G. Guiochon, P.K. Dasgupta, R.A. Shalliker, Anomalies in evaporative light scattering detection, *Analytica Chimica Acta*, 703 (2011) 245-249.
- [25] E.T. Whittaker, XVIII.—On the Functions which are represented by the Expansions of the Interpolation-Theory, *Proceedings of the Royal Society of Edinburgh*, 35 (1915) 181-194.
- [26] H.D. Luke, The origins of the sampling theorem, *Communications Magazine, IEEE*, 37 (1999) 106-108.
- [27] C.E. Shannon, Communication in the Presence of Noise, *Proceedings of the IRE*, 37 (1949) 10-21.

- [28] Luciano da Fona Costa, J. Roberto Marcond Cesar, Shape Classification and Analysis: Theory and Practice, 2nd ed., CRC Press, Taylor and Francis Group, Boca Raton, FL, 2001.
- [29] R.G. Mortimer, Mathematics for Physical Chemistry, 4th ed., Academic Press, Amsterdam, 2013.
- [30] Y. Vanderheyden, K. Broeckhoven, G. Desmet, Comparison and optimization of different peak integration methods to determine the variance of unretained and extra-column peaks, *Journal of Chromatography A*, 1364 (2014) 140-150.
- [31] R.G. Lyons, Understanding Digital Signal Processing, 3rd ed., Prentice Hall 2010.
- [32] D. Schlichthärle, Digital Filters: Basics and Design, 2 ed., Springer Verlag, Berlin Heidelberg, 2011.
- [33] K.J. Fountain, U.D. Neue, E.S. Grumbach, D.M. Diehl, Effects of extra-column band spreading, liquid chromatography system operating pressure, and column temperature on the performance of sub-2- μ m porous particles, *Journal of Chromatography A*, 1216 (2009) 5979-5988.
- [34] Kraiczek, K., Agilent Technologies, Germany. Personal communication July 2015
- [35] Method E 1657: Standard Practice for Testing Variable-Wavelength Photometric Detectors Used in Liquid Chromatography
- [36] C.G.E. Howard V. Malmstadt, Stanley R. Crouch, Microcomputers and Electronic Instrumentation-Making the Right Connections, American Chemical Society, Washington DC, 1994.
- [37] P.K. Dasgupta, Chromatographic peak resolution using Microsoft Excel Solver: The merit of time shifting input arrays, *Journal of Chromatography A*, 1213 (2008) 50-55.
- [38] A. Kadjo, P.K. Dasgupta, Tutorial: Simulating chromatography with Microsoft excel macros, *Analytica Chimica Acta*, 773 (2013) 1-8.
- [39] F.X. Diebold, Elements of Forecasting, 4 ed., Thomson Higher Education, Mason, 2006.
- [40] Texas Instruments; Direct RF-Sampling ADCs Sample Beyond 3.0 GHz, <http://www.ti.com/lscs/ti/data-converters/high-speed-adc-greater-than-1-gsps-rf-sampling.page>
- [41] J.C. Russ, The Image Processing Handbook, CRC Press, Taylor and Francis Group, Boca Raton, FL, 2011.
- [42] L. M. Kegelmeyer, P.W. Fonga , S.M. Glenna, J. A. Liebmana, Local Area Signal-to-Noise Ratio (LASNR) algorithm for Image Segmentation, http://ai.stanford.edu/~fongpwf/LASNR11_spie_6696-85-o.pdf

- [43] D. Bissieres, I. Campbell, Y. Lesperat, Analog to Digital Converter, http://ume.gatech.edu/mechatronics_course/ADC_F04.ppt (accessed August 2015)
- [44] S. Butterworth, On the theory of filter amplifiers, *Wireless Engineer*, 7 (1930) 536-541.
- [45] Maxim Integrated, Tutorial 733, Filter Primer, <http://www.maximintegrated.com/en/app-notes/index.mvp/id/733> (accessed July 2015)
- [46] S.W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*; Chapter 15, Moving Average Filter, 1st edition ed., California Technical Pub San Diego, California, 1997.
- [47] G.J. Butterworth, Effect of Detector Response Time on Distortion of Spectral Peaks of Gaussian Shape, *Journal of Physics E-Scientific Instruments*, 1 (1968) 1165-&.
- [48] J.V. Hinshaw, Handling fast peaks, *LC-GC North America*, 21 (2003) 268.
- [49] R.A. Shalliker, S. Kayillo, G.R. Dennis, Optimizing Chromatographic Separation: An Experiment Using an HPLC Simulator, *Journal of Chemical Education*, 85 (2008) 1265.
- [50] V.R. Meyer, *Practical High-Performance Liquid Chromatography*, Wiley, West Sussex, England, 1994.
- [51] M. Goedert, G. Guiochon, Sources of error in chromatographic analysis. Effect of sampling parameters on the accuracy of numerical measurements, *Chromatographia*, 6 (1973) 76-83.
- [52] K.L. Rowlen, K.A. Duell, J.P. Avery, J.W. Birks, Whole column detection: application to high-performance liquid chromatography, *Analytical Chemistry*, 61 (1989) 2624-2630.
- [53] J.M. Anderson, Data Acquisition and Processing for High-Speed Liquid-Chromatography, *Journal of Liquid Chromatography*, 6 (1983) 2809-2828.

Figure Captions

Figure 1: Determination of frequency components in narrow Gaussian peaks by Fast Fourier transform (See experimental for the software details). (A) The Gaussian peaks are centered on zero for clarity. The standard deviations in time domain are as follows: blue =0.08 s, red =0.2 s and green 1.5 s. (B) FFT of the peaks in the frequency domain

Figure 2: Association of response or rise times with a digital filter. The effect of multi-pass centered 5-point moving average filter is shown on a unit step function. (See supplementary files for other types of filters used in modern chromatography)

Figure 3. Simulated ultrafast chromatogram showing the effect of sampling frequency (160, 40, 4 and 2.5 Hz) without any electronic effect on four peaks of different height, noise and resolution. White Gaussian noise has been added. The inset shows the noise level with sampling frequency. The efficiencies are set as follows for peak (1) 15,529, (2) 14,135 (3) 13,751, (4) 13,830 corresponding to standard deviations in the time domain σ_t of 0.13, 0.18, 0.19 and 0.25 s respectively. The y-axis is same for all chromatograms

Figure 4. Details of the first peak in Figure 3 as it would appear at different sampling frequencies. There is no data averaging except for the dashed line and all plots have the same ordinate scaling although they have been vertically displaced for clarity. The dashed red line shows what a real 40 Hz output may look like where every successive four data points have been averaged in an original 160 Hz sample. Note 2-fold decrease in noise.

Figure 5. Effect of sampling frequency on the observed efficiency and noise level in high efficiency chromatography. System 3, time constant fixed at 0.01s. The first peak is uracil and second peak is that of phenol. Injection volume 1 μ L, mobile phase: 80:20 ACN:H₂O at 1.8 mL/min, detection wavelength 254 nm. The peak efficiency was calculated assuming Gaussian peaks. Standard deviation (SD) are shown for the phenol peak

Figure 6. Convolution of 1 Hz square wave by digital filtering on two different liquid chromatographs; System 3 (A to D), and System 1 1260 (E to H). The system 3 software uses “time constant” and system 1 uses a “response time”. See supplementary information for the names of the instruments. Wavelength of detection is 610 nm.

Figure 7. Effect of the choice of the time constant on the observed efficiency and noise level. System 3 with a fixed sampling frequency of 100Hz. The first peak is uracil and second peak is that of phenol. Injection volume 1 μ L, mobile phase: 80:20 ACN:H₂O at 1.8 mL/min, detection wavelength 254 nm. Standard deviations are shown for the phenol peak.

Figure 8. Effect of coupled digital filtering and sampling frequency on the observed efficiency and noise level in high efficiency chromatography. System 1. The first peak is uracil and second peak is that of phenol. Injection volume 1 μ L, mobile phase: 80:20 ACN:H₂O at 1.8 mL/min, detection wavelength 254 nm. Standard deviations (SD) are shown for the phenol peak.

Table 1. Analysis of the suggestions given by various authors on the sampling frequency for chromatographic purposes

Data Sampling Suggestion	Comments	Ref
20 data points per peak	Suggestion as “generally accepted rule” for the smallest peak in the chromatogram	[20]
100 data points per peak	Moment analysis; ≥ 100 points/peak are suggested for accurate peak areas	[16]
$\frac{\sigma_G}{t_s} > \frac{n}{\pi}$; The ratio $\frac{\sigma_G}{t_s}$, determines the minimum number of sampling points	Fourier analysis; t_s is sampling interval (s), σ_G = standard deviation of the narrowest Gaussian peak, n is harmonic component of the Fourier series. It is suggested that “time constant” should be 1/3 of sampling period.	[51]
For Gaussian function 6-7 points are sufficient to describe it	Calculates 7 points per peak for less than 0.1% error in peak area (uses Shannon’s theorem)	[18]
10 data points across $2.355 \sigma_G$	Provides a "good guideline" for fast peaks	[48]
8 data points across baseline width of a Gaussian $8\sigma_G$	No criterion provided	[52]
36 data points per peak	Approximation derived from the inverse relationship between standard deviation of the Gaussian function and its Fourier transform	[17]
14 data points per peak	Analysis based on allowed maximum error (0.001%) in the peak height	[14]
sampling frequency $= \frac{CQ\sqrt{N}}{4V_o(1+k)}$	C= no. of data points desired per peak, Q= volumetric flow rate, N = plates, V_o = void volume, k= retention factor	[53]
Real time oversampling (up to MHz) and data averaging	Shows the advantage of oversampling points into one recorded data and averaging to improve the signal to noise ratio	[19]
15 data points for a Gaussian peak	15 data points for < 0.1 % error in peak area; Simpson’s rule: 6 x base width	[15]
Data acquisition rate has no influence on band broadening and resolution	Also shows an example of "hidden features" in the software causing peak distortion	[21]

Table 2. Effect of RC type filtering on chromatographic peak characteristics on an initial Gaussian peak*

$\tau/w_{0.5}$	10-90% response time/ $w_{0.5}$ *	increase in $w_{0.5}$ (%)	decrease in (N) (%)	decrease in peak height (%)	asymmetry at 5% peak height	asymmetry at 10% peak height
0.0231	0.0507	0.1155	0.2308	0.1470	1.0007	1.0004
0.0462	0.1015	0.5389	1.0748	0.5767	1.0022	1.0017
0.0693	0.1522	1.2317	2.4483	1.2597	1.0066	1.0059
0.0924	0.2030	2.1170	4.1892	2.1559	1.0142	1.0121
0.1155	0.2537	3.1563	6.2129	3.2228	1.0261	1.0215
0.1386	0.3044	4.3495	8.5098	4.4208	1.0417	1.0344
0.1617	0.3552	5.6197	10.9236	5.7153	1.0607	1.0500
0.1848	0.4059	6.9669	13.4484	7.0774	1.0834	1.0687
0.2079	0.4567	8.3911	16.0780	8.4838	1.1089	1.0886
0.2309	0.5074	9.8152	18.6671	9.9158	1.1376	1.1111
0.3464	0.7611	17.4365	31.8327	17.0504	1.3150	1.2502
0.4619	1.0148	25.2887	44.1822	23.6305	1.5276	1.4192
0.5774	1.2685	33.2948	55.5042	29.4507	1.7523	1.6029
0.6928	1.5222	41.3010	65.5443	34.5416	1.9829	1.7938
0.8083	1.7759	49.4226	74.4193	38.9930	2.2146	1.9873

* See Supplementary Information Excel macros for more information

Fourier Analysis of Narrow Gaussian Peaks

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6

Figure 7

(A) 80 Hz, 0.063 s

(B) 40 Hz, 0.13 s

(C) 10 Hz, 0.5 s

(D) 1.25 Hz, 4.0 s

Figure 8

M. Farooq Wahab

M. Farooq Wahab is a visiting scholar in Prof. Armstrong's start-up company at the University of Texas at Arlington. He completed his PhD from the University of Alberta on FS Chia Scholarship where his research focused on the development of novel high efficiency phases. He has received several academic awards from the University of Karachi and University of Alberta. His interests include fundamentals of chromatography, with an emphasis on developing new phases for hydrophilic molecules and understanding suspension flow properties for packing ultrahigh efficiency particles. He also developed low cost pedagogical spectroscopy experiments, which have been adapted by many universities.

Daniel W. Armstrong

Daniel W. Armstrong is the Welch Distinguished Professor of Chemistry at the University of Texas at Arlington. Professor Armstrong has received over 30 national, international research and teaching awards. His current interest involves chiral recognition, macrocycle chemistry, synthesis and use of ionic liquids, separation science and mass spectrometry. He has over 600 publications, one book and 30 patents. He has founded two companies, was on the scientific advisory board of three corporations and one university. His work provided impetus for the FDA's regulatory changes regarding chiral drug development in 1992. He developed the most comprehensive solvation and characterization models for ionic liquids.

Purnendu K. (Sandy) Dasgupta

Purnendu K. (Sandy) Dasgupta is the Hamish Small Chair in Ion Analysis at the Department of Chemistry and Biochemistry at the University of Texas at Arlington, U.S.A. Liquid/ion chromatography and electronics in instrumentation are among his research interests. He received the American Chemical Society Award in Chromatography in 2011, the J. Calvin Giddings Award in Chemical Education in 2015, and was also elected a Fellow of the Institute for Electrical and Electronics Engineers in 2015. He has authored 400+ publications and has been one of the editors of *Analytica Chimica Acta* since 2004.

.

Akinde Florence Kadjo

Akinde Florence Kadjo is a third year graduate student at the University of Texas at Arlington. She is originally from the Ivory Coast. She has mainly worked on excel simulations and cyanide detection in blood.

Consequences of Detector Settings on Ultrafast and Ultrahigh Efficiency Chromatography