Utah State University

From the SelectedWorks of Stephen E. Bialkowski

January 1, 1987

Optimal Estimation of Impulse-Response Signals Through Digital Innovations and Matched FilterSmoothing

Stephen E. Bialkowski, *Utah State University*

Available at: https://works.bepress.com/stephen_bialkowski/22/

AIP | Review of Scientific Instruments

Estimation of impulseresponse signals through empirically derived digital matched filter smoothing

Stephen E. Bialkowski

Citation: Review of Scientific Instruments **58**, 687 (1987); doi: 10.1063/1.1139239
View online: http://dx.doi.org/10.1063/1.1139239
View Table of Contents: http://scitation.aip.org/content/aip/journal/rsi/58/4?ver=pdfcov
Published by the AIP Publishing

Articles you may be interested in

Estimation of vocaltract area function from lip impulse response based on lossy vocaltract digital filter model
J. Acoust. Soc. Am. **110**, 2776 (2001); 10.1121/1.4777742

Photopyroelectric thinfilm instrumentation and impulseresponse detection. Part III: Performance and signal recovery techniques
Rev. Sci. Instrum. **58**, 2033 (1987); 10.1063/1.1139511

Stochastic characterization of propagation and target impulseresponse measurements
J. Acoust. Soc. Am. **60**, S56 (1976); 10.1121/1.2003420

ImpulseResponse Function of Periodically Supported Systems
J. Acoust. Soc. Am. **51**, 88 (1972); 10.1121/1.1981668

Frequency Spreading of Scattered Sound through ImpulseResponse Measurements
J. Acoust. Soc. Am. **50**, 133 (1971); 10.1121/1.1977631

AIP Publishing Re-register for Table of Content Alerts
Create a profile. Sign up today!

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitationnew.aip.org/termsconditions. Downloaded to IP: 129.123.125.251 On: Tue, 20 May 2014 17:40:18

Estimation of impulse-response signals through empirically derived digital matched filter smoothing

Stephen E. Bialkowski

Department of Chemistry and Biochemistry, Utah State University, Logan, Utah 84322-0300

(Received 15 September 1986; accepted for publication 12 December 1986)

A real-time digital filter is described which may be most useful for optimal determination of the magnitude of impulse-response functions found in pulsed, repetitive experiments of low duty cycle. This filter is based on a matched filter but employs an interference orthogonalization step. This results in a signal magnitude estimate which is independent of coherent interference. The filter updates the signal magnitude estimate upon each repetition of the experimental cycle. Comparisons to signal estimation using gated sampling devices are given.

INTRODUCTION

Of primary importance in modern chemical analysis is the estimation of analyte levels based on the magnitudes of instrumental responses.[1,2] The precision to which the analyte level can be reported is directly proportional to the precision to which the signal estimate can be obtained and reproduced.[3] Signal estimation can be simple, for example when the signal is a smooth function of time and the signal level is much greater than that of the noise. There are, however, several instrumental methods where the analytical signal is a measured physical response to an often periodic impulse excitation of the analyte system. In these cases, the analytical signal can be decomposed into two primary components. The first component is the temporal response to the excitation. This component is the impulse-response function in cases where the excitation is an impulse, or is the convolution of the impulse-response with the excitation when the excitation is longer than the response time of the system. The impulse-response function typically contains information relevant to the analyte system under study and the physics particular to the instrumental method being used. Thus the impulse-response function can be thought of as a qualitative description of the system. The second component of the signal is that of the magnitude of this impulse-response function. This component contains information on the amount of analyte in the system and is thus a quantitative estimate. It is often the case that the impulse-response function is known and only an estimate of the magnitude is desired. The latter occurs when the impulse-response function is not a qualitative description, but rather is due to the physics of the particular instrumentation.

There are several instruments currently employed to sample and thereby estimate transient periodic data. The two most common are the gated integrator, or boxcar averager, and the transient recorder.[4,5] For the most part, the gated integrator is used when the magnitude of the signal is to be estimated. For example, it is common to find sample and hold circuits, a simple form of the gated integrator, used in the estimation of signal magnitudes in pulsed laser excited spectroscopic methods of analysis.[6–8] These circuits are used to smooth the low duty cycle periodic signals so that slow response time instrumentation can be used to record these data. The signal itself is a cyclostationary process in that the transient occurs periodically and is reproduced over each cycle of the experiment.[9] Each time an experimental cycle occurs, the signal is sampled at some fixed delay relative to the start of the cycle resulting in replicate measurement of the signal magnitude. For zero mean noise, independent of the signal, the accuracy of the signal estimate increases as the square root of the number of replicate measurements obtained. Estimation of the precision of the signal estimate also becomes more accurate as more and more replicate measurements are taken. However, this increase in estimation accuracy is obtained only if the signal magnitude and impulse-response change on time scales long compared to the experimental cycle time. If the signal magnitude is dynamic, averaging over several cycles is not valid. In this case postexperimental smoothing is often employed.

The transient recorder acts as a number of sample and hold circuits, each sampling the transient signal at a different time delay relative to the start of the cycle. This instrument can be used for multichannel averaging if the appropriate software or hardware is used.[5] In this case, the impulse-response function and the magnitude of this function are both obtained. The accuracy of estimation of these parameters again increases as the square root of the number of averaged transients. As in the case of gated averaging, the multichannel averaging procedure is valid only for samples that do not change over the analysis time. Signal estimation applications using the transient recorder to capture signal transients that change with the experimental cycle are not as common as that utilizing the sample and hold circuits. However, this type of estimation can be performed using more complex sampling functions. And there are, in fact, several advantages to the latter method over that of the simple sample and hold. The advantages of this application are discussed below.

In this laboratory, a transient recorder has been used to capture data for both multichannel averaging and to perform the mathematical equivalent of gated sampling.[10–14] Recently, a correlation technique derived from a least-squares analysis of the signal has been developed and utilized

 0034-6748/87/040687-09$01.30 © 1987 American Institute of Physics

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitationnew.aip.org/termsconditions. Downloaded to IP: 129.123.125.251 On: Tue, 20 May 2014 17:40:18

to recover signal magnitude estimates from low signal-to-noise ratio transients.[15] The latter is a matched filter technique in that the function used to filter the transient data is the expected signal.[9] Technically, it is a smoothing function in the sense that one value of the estimate is obtained for the entire experiment cycle. More recently, this type of smoothing has been extended to include signal estimation in the presence of interference which is synchronous to the experimental cycle.[16] The signal estimate obtained from this digital smoothing procedure can be independent of such interference if this interference is reproduced on each cycle. The latter procedure is a Wiener filter with prior innovations filtering.[9]

In this paper the detailed mathematical foundation of the digital smoothing filters used in the experimental works mentioned above will be presented. A method for determining the signal to noise of the estimate based on the particular sampling function utilized is illustrated. This paper does not discuss smoothing in the sense of data being reduced after the signal estimates have been obtained. This type of smoothing has been performed by apodization of the transformed data, as in frequency bandwidth limited processing,[9] procedures such as the Savitzky–Golay smoothing filter,[17] or similar but adaptive filters,[18] by Kalman filtering[19,20] and by matched and adaptive filtering.[21] These postexperimental smoothing procedures are sensitive to data acquisition and processing errors.[20] But there is often no attempt to quantify or reduce these errors. In contrast to these postexperimental smoothing filters, this paper addresses the real-time optimal estimate of the transient signal based on correlation filtering of the signal transient itself. This real-time matched filtering results in the optimum estimate of the signal by optimization of the signal-to-noise ratio of the predicted signals.

I. THEORY

A process is cyclostationary if its statistics are invariant to a shift in the time origin equal to integral multiples of the cycle time or period T.[9] The signal estimators described here are smoothing filters where the signal estimate is updated after each cycle. A schematic representation of the most general filter is illustrated in Fig. 1. The real input $x(t)$ is comprised of two main components,

$$x(t) = s(t) + v(t), \qquad (1)$$

with $s(t)$ being the analytical signal and $v(t)$ being the sum of uncorrelated noise and periodic coherent correlated noise or interference. The time t is that relative to the start of the current cycle nT. The time-dependent input has an associated finite transform $X(w)$, which is also a sum of two terms $S(w)$ and $V(w)$. These transforms are, in general, complex being obtained from real functions.[22] The uncorrelated noise is stationary in that the noise statistics do not vary over the cycle. Coherent noise is cyclostationary in that the noise statistics vary over the cycle and are reproduced on each cycle. Coherent noise is, in general, correlated to the cycle and may vary in time within the experimental cycle.

The first stage of smoothing is the whitening filter $F(w)$, with an impulse-response of $f(t)$. The output of this stage is the innovations

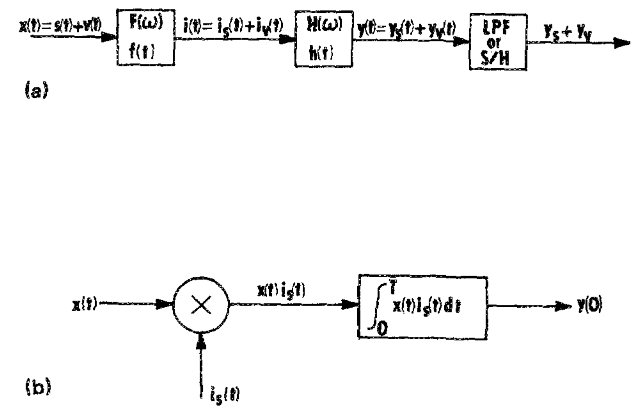

FIG. 1. A schematic representation of the real-time matched filter in both the frequency- and time-domain implementations. The filter (a) is composed of three stages: the whitening filter $F(w)$ with impulse-response $f(t)$, the matched filter $H(w)$ and $f(t)$, and the low-pass or sample and hold output filter. The input $x(t)$ composed of signal $s(t)$ and noise $v(t)$ is first "whitened" resulting in the innovations $i(t)$. $i(t)$ is then filtered resulting in $y(t)$. The output is sampled at the time of the maximum signal-to-noise ratio. The filter process (b) shows the implementation of (a) for a periodic input with prior knowledge of $s(t)$ and $n(t)$. The output is updated in units of T, the cycle period.

$$i(t) = i_s(t) + i_v(t), \qquad (2)$$

where $i(t)$ is composed of mutually orthogonal signal $i_s(t)$ and noise $i_v(t)$ components. This stage of the filter is such that nonwhite noise statistics of $x(t)$ result in white noise in $i(t)$. Due to the orthogonality of its components, the signal component of the innovations, can also be made independent of cyclostationary coherent noise in $x(t)$.[23] In this case $i_s(t)$ may be made orthogonal to $v(t)$. The orthogonal signal $i_s(t)$ is then used to estimate the signal magnitude, independent of the coherent noise, by the matched filter discussed below.

The second stage of the smoothing filter is the signal estimator $H(w)$ with an associated transform impulse-response function $h(t)$. The output of this state is the time-dependent signal estimate

$$y(t) = y_s(t) + y_v(t), \qquad (3)$$

which, again, is made up of two components: that due to the signal $y_s(t)$ and that due to the noise $y_v(t)$. The output of this filter is a function of the time in the cycle. The signal magnitude estimate at a time in the cycle when the signal-to-noise power ratio (SNR) is a maximum is desired. This estimate is obtained by sampling the output of the first two stages of the filter at a time relative to the start of the cycle. In applications where the signal magnitude is expected to vary smoothly from cycle to cycle, a low-pass filter may be used in place of the sample and hold.

The simplest implementation of the filter occurs for nonperiodic, uncorrelated white noise. In this case the whitening filter does not modify the input $x(t)$. This is the matched filter. The frequency domain filter $H(w)$ is, in general, a complex function composed of a real magnitude and an imaginary phase since it is the transform of the real impulse-response filter $h(t)$ which is generally one sided, asymmetric about the arbitrary time origin. Because $H(w)$ is typi-

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitationnew.aip.org/termsconditions. Downloaded to IP: 129.123.125.251 On: Tue, 20 May 2014 17:40:18

cally complex, frequency domain matched filtering is limited. The time and frequency operations of the filter are related by Parseval's theorem. The noise power at the output of the filter at some time t is

$$y_v^2(t) = (2\pi)^{-2} \int_{-\infty}^{\infty} |V(w)H(w)|^2 dw. \quad (4)$$

For white noise, $V(w)^2/2\pi$ is a constant σ^2 and thus

$$y_v^2(t) = (2\pi)^{-2}\sigma^2 \int_{-\infty}^{\infty} |H(w)|^2 dw. \quad (5)$$

The signal power at the output is the squared frequency filtered signal

$$y_s^2(t) = (2\pi)^{-2} \left| \int_{-\infty}^{\infty} S(w)H(w)e^{iwt} dw \right|^2. \quad (6)$$

The SNR is the ratio of the last two equations. If $x(t)$ is either a current or a potential signal,

$$\mathrm{SNR} = y_s^2(t)/y_v^2(t) = \left| \int_{-\infty}^{\infty} S(w)H(w)e^{iwt} dw \right|^2 \Big/ 2\pi\sigma^2 \int_{-\infty}^{\infty} |H(w)|^2 dw. \quad (7)$$

The infinite integration limits are used for mathematical simplicity and are never realized. The frequency interval will be limited to that of the instrumental response or the Nyquist frequency in the case of sampled signals. In many cases it is easier to use time domain equivalent found from Parseval's theorem, thus avoiding this problem.[22]

The matched filter is formulated in such a fashion as to maximize the SNR of the resulting signal estimate. Using Eq. (6) and Schwartz' inequality yields

$$y_s^2(t) = (2\pi)^{-2} \left| \int_{-\infty}^{\infty} S(w)H(w)e^{iwt} dw \right|^2 \leqslant (2\pi)^{-2} \int_{-\infty}^{\infty} |S(w)|^2 dw \int_{-\infty}^{\infty} |H(w)|^2 dw, \quad (8)$$

where the equality is the maximum value. Equation (8), and thus the SNR, is a maximum if and only if the filter function is to within an arbitrary constant of the complex conjugate of the signal,[9,21]

$$H(w) = kS^*(w)e^{-iwt_0}, \quad (9)$$

where k is an arbitrary constant, $S^*(w)$ is the complex conjugate of the signal transform and the exponent is an arbitrary phase constant. The exponential phase factor may be set equal to unity when the experimental cycle initiation time is known since t_0 can be arbitrarily chosen to be zero. Frequency-domain filtering has the corresponding time-domain process found from the inverse transform of the process

$$y_s(t) = k \int_{-\infty}^{\infty} S^*(w)X(w) dw = k \int_0^{\infty} s(t')x(t+t')dt', \quad (10)$$

where $y_s(t)$ is the signal estimate at time t. The fact that the optimum SNR is found for the correlation of the signal $x(t)$ with the expected signal $s(t)$ results in this filter process being known as matched filtering. The correlation integral over all retardation times t will result in the maximum correlation at the synchronous experimental cycle start time and subsequently the entire correlation is not required to obtain the optimum signal estimate. The time-dependent signal estimate obtained from the correlation in Eq. (10) can be simplified when the experimental cycle is synchronous to some event indicating the cycle. In this case only one correlation integral need be calculated[15,16]:

$$y_s = k \int_0^{\infty} s(t)x(t)dt. \quad (11)$$

Treatment of the filter in the event of nonwhite noise stationary statistics is straightforward in the systems approach.[9,22] In this case the optimum SNR is obtained for the signal after the operation of the whitening filter,

$$y_s = k \int_0^{\infty} i_s(t)i(t)dt, \quad (12)$$

where $i_s(t)$ is the impulse-response of the frequency domain filter function H(w) given by[9]

$$H(w) = kS^*(w)e^{-iwt_0}/S_v(w) \quad (13)$$

and $S_v(w)$ is the noise power spectrum obtained from the transform of the noise autocorrelation. In this formulation the filter function includes both the whitening filter and the matched filter functions. Equation (13) is valid for stationary noise statistics since the noise power spectrum is a real function. However, for a cyclostationary process, $S_v(w)$ will be complex since noise statistics are time dependent. It is easier to treat the filter stages independently, thereby avoiding the problem associated with the cyclostationary noise statistics. In this case $S^*(w)/S_v(w)$ is replaced by the transform of the innovations $I_s^*(w)$. Since $i(t)$ is equivalent to $x(t)$ in all respects except for the basis set used to construct the particular functions, $x(t)$ can be substituted for $i(t)$ in Eq. (12).

The above discussions are based upon continuous functions. However, the recorded and subsequently digitized signals are a series of discrete estimations. Descriptions of continuous function statistics are related to the discrete functions through sampling theory.[9,22,24] In general, the results obtained for discrete function statistics are the same as those of the continuous function with a few considerations. The main consideration is that of the Nyquist limited frequency interval and the bandpass associated with the finite sampling time of the individual gates. With attention paid to these band-limiting processes, it can be shown that the reconstructed sampled function is identical to that of the sampled continuous function,[22] and so mathematical treatment of discrete functions with continuous function mathematics is a valid endeavor.

II. EXPERIMENTAL

The data discussed in this paper were obtained using either a photothermal lens or a photothermal deflection spectroscopy apparatus. Both of these apparatuses have been previously described in detail elsewhere.[10–16] In both apparatuses, a pulsed laser is used as the impulse source of sample excitation. The signal is derived from a deviation in the light path of a second continuous wave laser used to monitor the photothermal refractive index perturbation

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitationnew.aip.org/termsconditions. Downloaded to IP: 129.123.125.251 On: Tue, 20 May 2014 17:40:18

caused by the impulse excitation laser. The excitation laser pulse widths used in these experiments ranged from 5 ns, for the molecular nitrogen laser, to 170 ns for the molecular carbon dioxide laser. A silicon photodiode in conjunction with appropriate spatial filters monitors the effect of the perturbation on the continuous laser beam intensity or position. Chemical samples are introduced into the apparatus using either a gas or liquid cell. The samples can be of either fixed composition or variable composition such as the effluent from a gas chromatograph.

Signals observed the silicon photodiode obtain maximum values during or shortly after (that is, within a few ns) the pulsed laser excitation. After the pulsed laser excitation, the signal decays with a functionality characteristic of thermal diffusion. The signal can, in theory, be described by $s(t) = s_0(1 + 2t/t_c)^{-2}$, where s_0 is the maximum signal at zero time and t_c is the characteristic time constant dependent on the thermal diffusivity and the optical beam waist radius. The maximum signal s_0 is a function of the analyte absorbance and the excitation laser pulse energy. Thus, the functional characteristic of the signal is a product of a magnitude term with that of an impulse-response function. For dilute samples of less than about 1%, the impulse-response function is a characteristic of the apparatus and not the analyte.

Scaling of the signal magnitude for variations in the pulsed excitation laser intensity is not performed as in conventional spectrophotometry. The pulsed laser energy is monitored with a power meter and the output of this meter is recorded by the DEC LSI 11/23 processor used in this laboratory. These laser energy data are stored in a data array for future use in examining the correlation with the signal magnitude estimate. The signal itself is a modulation of the continuous laser beam intensity. The analytical signal current from the silicon photodiode is generally followed with a transimpedance amplifier, and the subsequent time-dependent voltage is recorded with a Physical Data Incorporated model 522A transient waveform recorder. This 20-MHz, 8-bit transient recorder utilizes a TRW flash analog-to-digital converter, which in turn has an effective sampling aperture duration of a few ns. The bandwidth of this device is limited by the input amplifier set to 3 dB at 10 MHz. The transient recorder is interfaced to the processor with a parallel line interface unit which is operated under software control.

A main program, run under DEC RT-11 real-time operating system, serves to sequence the operations of multichannel averaging, innovations construction, and digital filtering. The main program is written in FORTRAN to facilitate the interaction between user and apparatus, and also to facilitate the interface with the DEC SSP scientific software subroutine package used for regression and correlation analysis. The individual subroutines which access the data recording devices are written in MACRO. For example, multichannel averaging, and digital matched filtering routines are both written in MACRO. The MACRO language is preferred when speed is important. The main program first calls subroutines which obtain multichannel averaged coherent noise and signal plus coherent noise data vectors. A signal vector is obtained from the difference of these two. The innovations vector is then obtained with a modified, in place, Gram–Schmidt orthonormalization routine.[23] Up to three basis set vectors are used. The first vector is a constant, base-line vector, and is always used. Inclusion of this vector ensures that the signal estimate will not be a function of baseline drift. The second vector may be the coherent noise or interference data. This vector is often omitted if the coherent noise is, for practical purposes, a constant baseline term. The last vector is always that of the expected signal.

After the Gram–Schmidt computations are complete, the last vector is multiplied by the last term of the inverse factorization matrix, as discussed below. This vector is then the optimum filter for the expected signal as determined for the particular experiment being performed.

Once the optimal filter has been formulated, the actual experiment is run. Experiments typically run in this laboratory include those where the pulsed laser energy dependence of the signal magnitude is sought and where the signal magnitude will change on each experimental cycle initiated by the laser pulse, as in the case of chromatographic effluent analysis. In all cases, a signal magnitude estimate is obtained and stored in a data array. Each consecutive point in this array corresponds to a consecutive experiment cycle $s(nT)$.

The start of the experimental cycle is synchronous to the power line. The repetition rate was 3.75 Hz. A pulse derived from the line frequency initiates the pulsed laser, pulsed laser energy recording, and triggers the transient digitizer. After a transient signal is recorded, the 8-bit data is transferred, word by word, to the processor. Each word is loaded into a floating point register, converted to floating point format, and multiplied by a floating point format filter element that is supplied from the main routine as a time-ordered array. The resulting product is summed into a floating point accumulator. After all data have been transferred and the floating point computations performed, the accumulator contains the dot product of the filter function with the data. This value is the signal estimate and is returned to the main routine or stored in an array. At the end of an experiment consisting of n cycles, all data are written to disk for later analysis and archival storage. The results reported on below are from a collection of such data.

III. RESULTS

A. Synchronous experimental cycles and multichannel averaging

It is commonly believed that the experimental cycle frequency should not be an integer multiple or subdivision of the power line frequency if accurate estimations are to be obtained with transient digitizers.[4] This rule is based on the fact that for asynchronous cycle frequencies, line interference will be effectively independent of the signal and thus will not add coherently to the averaged signal. On the other hand, if the synchronous interferences are reproducible in their frequency distribution and phase, but not necessarily in magnitude, then more efficient signal estimation can be obtained by innovations filtering. In this case the experimental cycle must be synchronous to the line frequency. The inter-

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitationnew.aip.org/termsconditions. Downloaded to IP: 129.123.125.251 On: Tue, 20 May 2014 17:40:18

ference is eliminated by the orthogonalization, an effective notch filter. Often the synchronous coherent interference does not originate from an independent source such as power line radiation but is an intrinsic part of the experimental excitation itself. In this instance, the experimental cycle cannot be made asynchronous with the interference and a corrective process should be performed in order to obtain accurate results.

A common method for obtaining interference independence involves obtaining multichannel averaged signal and background transients, followed by subtraction of the background from the signal.[4,5] The measured signal is comprised of three independent terms,

$$x(t) = s(t) + v_c(t) + v_i(t), \tag{14}$$

where $v(t)$ of Eq. (1) has been decomposed into $v_c(t)$, the coherent noise or interference, and $v_i(t)$ is the independent stationary noise. The analytical signal $s(t)$ will only be present in the signal when an analyte is present in the measurement apparatus and is being stimulated by the excitation source. Since $v_i(t)$ is independent of $s(t)$ and $v_c(t)$, multichannel signal averaging of the analyte signal will result in an increase in the SNR proportional to the number of transients averaged and ultimately resulting in the determination of $x(t) = s(t) + v_c(t)$. The important vector is that of the signal $s(t)$. The latter is determined as the difference of the two averaged transients.

Figure 2 illustrates a typical noise power spectrum obtained for an experimental apparatus designed to measure photothermal deflection by an increase in the amount of light of a probe laser on a *pin* photodiode detector.[10,13,14] The main single source of noise in these data are the interference due to power line pickup at 60 Hz. Consider the case where the data are to be obtained with a Nyquist limit of 1 kHz. If upon each experimental cycle an estimate of the signal is obtained and summed into a composite estimate then the total signal estimate will increase proportional to the number of individual estimates,[25]

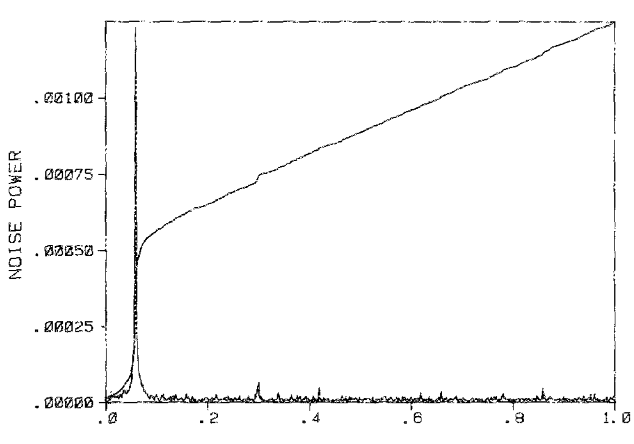

FIG. 2. Curve A is a noise power spectrum obtained from one average of the experimental apparatus. The noise power is in squared volts. Curve B is the relative numerical integral of this noise power. Notice that about half of the noise power over this limited band is due to the line interference at 60 Hz.

$$s_{\rm tot} = Ms, \tag{15}$$

where M is the number of signal estimates or experimental cycles, s is the signal, and $s_{\rm tot}$ is the total estimate. On the other hand, the variance due to the zero mean incoherent noise, $v_i(t)$ will increase proportional to M[9,24]

$$\sigma^2_{\rm tot} = M\sigma^2 = M(2\pi)^{-1}\int_{-\infty}^{\infty} |V_i(w)|^2 dw, \tag{16}$$

where $\sigma^2_{\rm tot}$ is the total noise power after signal averaging, σ^2 is the noise power of one experimental cycle, and $V_i(w)$ is the transform of the independent noise. The variance for a single estimate is, in turn, proportional to the noise power. The SNR of the signal power estimate is

$$\text{SNR} = s^2_{\rm tot}/\sigma^2_{\rm tot} = Ms^2/\sigma^2, \tag{17}$$

thereby resulting in the well-known fact that the SNR improvement of the voltage or current is proportional to the square root of the number of estimates.

In the case of line-synchronous signal averaging, the signal estimate is the difference between the signal plus interference and the interference only data. The total variance is thus the sum of the variances of each of the two data. Thus $\sigma^2_{\rm tot} = 2M\sigma^2$. However, the interference is coherently averaged and eliminated from the final signal estimate by the difference technique. From the slope of the noise power data illustrated in Fig. 2, the total noise power over the 1-kHz band and independent of the interference is 5 nW in a 1-MΩ load. On the other hand, and assuming zero mean valued line interference, asynchronous signal averaging for $2M$ experimental cycles with a 1-kHz Nyquist limit will also result in a total variance of $2M\sigma^2$, but the total integrated noise power of Fig. 2 is 8.5 nW. Thus synchronous signal averaging results in an enhanced SNR over that of the asynchronous experimental method for an equivalent number of averages, in this case by a factor of about 1.3. In general, the SNR enhancement obtained by interference synchronous signal averaging is given by the ratio of the synchronous SNR to that of the asynchronous

$$\text{SNR}_{\rm synch}/\text{SNR}_{\rm asynch} = (\sigma^2 + \sigma^2_{\rm int})/\sigma^2, \tag{18}$$

which is greater than or equal to one. In the latter equation, σ^2 is the white-noise component of the total noise and $\sigma^2_{\rm int}$ is the interference component. Contrary to the belief that the experimental cycle should not be synchronous with the power line, Eq. 18 illustrates that there is in fact an advantage to performing synchronous multichannel signal averaging of periodic cyclostationary processes.

B. Innovation filter for known signal components

Background corrected multichannel averaging yields information of both the signal impulse-response function and its magnitude. If the impulse-response function is known, then the time-dependent data are a redundant estimation of the signal magnitude. However, these averaged data may be used to obtain the optimum SNR estimate of the signal magnitude. An optimal estimate of the magnitude can be obtained by matched filter smoothing of the averaged data using the expected signal [Eq. (11)]. It can be shown that this method is equivalent to least-squares fitting of the

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitationnew.aip.org/termsconditions. Downloaded to IP: 129.123.125.251 On: Tue, 20 May 2014 17:40:18

data to the expected signal[15] and to correlation between the data and the expected signal.[4]

The method of innovations construction utilizes the matched filter for signal estimation. In this method, an updated signal magnitude estimate is obtained after each experimental cycle. Subsequently, the method is more applicable to situations where either the analyte concentration is time dependent or the signal is nonlinear in its response to the excitation source. Innovations construction requires prior knowledge of the signal impulse-response function and the coherent noise functions. These functions can, in turn, either be obtained from multichannel averaged data or can be those expected from a theoretical standpoint.[15,16] In either event, the innovations are obtained by orthonormalization of the basis set of functions representing the expected signal and coherent noise contributions.

Orthonormalization can be performed using a Gram–Schmidt factorization procedure.[7,16,23] The square normalized basis functions for the expected signal $s_0(t)$ and coherent noise $v_{c0}(t)$ which may include more than one function, make up a column matrix **X** of rank m, where m is the number of basis vectors used to construct the matrix. The whitening filter is an m by m square, upper triangular, factorization matrix **F** used to form the innovations matrix **I**

$$\mathbf{I} = \mathbf{XF}. \tag{19}$$

The product of an input $x(t)$ with the innovations matrix results in coefficient vector **b** with elements corresponding to the weights of the orthonormal innovations vectors used to reconstruct the input

$$\mathbf{b} = x(t)\mathbf{I}. \tag{20}$$

The innovations filter **L**, defined as the inverse of the whitening filter $\mathbf{L} = \mathbf{F}^{-1}$ can be used to reconstruct the input in terms of the original basis set:

$$x(t) = \mathbf{LI}^{+}\mathbf{b}, \tag{21}$$

where $^+$ denotes the transpose.

The signal impulse-response $s(t)$ is known and the magnitude of this signal y_s is the only value of importance for signal estimation. Total reconstruction of the input $x(t)$ by the innovations filter is not required. The Gram–Schmidt procedure is a step-wise orthogonalization and normalization process. The first component of **X** is first normalized without changing the impulse-response function. Subsequent components of **X** are made orthonormal to the previous components. The basis set **X** is arranged such that the signal $s(t)$ is the last component in this matrix. The last component in the orthogonormalized matrix will be $i_s(t)$. $i_s(t)$ is orthogonal to the coherent noise components. More importantly, since the innovations filter matrix is upper triangular, the last coefficient of this matrix, $L_{m,m}$ is all that is required to scale the signal magnitude estimate to the correct value. Thus the coherent noise independent signal magnitude estimate is[16]

$$y_s = L_{m,m} \int_0^\infty x(t) i_s(t) dt, \tag{22}$$

where y_s is the optimal estimate of the signal magnitude since it results from matched filter smoothing with a whitening filter as illustrated in Fig. 1. More precisely, the basis set orthogonalizing whitening filter has been combined with the matched filter resulting in a linear filter system which is independent of coherent noise.

The implications of this filter method are important. First, the impulse-response function may be of any form. Determined experimentally, via multichannel averaging, the theoretical impulse-response function does not have to be known. Moreover, the theoretical impulse-response function may not be a useful filter basis because instrumental response times may significantly alter this function. The empirical basis obtained from multichannel averaging will result in more exact expected signal in this case. Second, if the experimental cycle is synchronous to coherent noise, and because of the particular ordering of the basis set used to construct the innovations vector through Gram–Schmidt orthonormalization, only one vector dot product is calculated for signal magnitude estimation. This allows for rapid estimation of the signal magnitude. Neglecting the time required for calculation of the innovations, this particular filter must be among the fastest digital filter techniques since no Fourier transforms are required and the full time correlation is not required because of the synchronous signal. And third, the signal estimate is independent of coherent interference. This feature can be obtained in frequency-domain filter methods. It is, however, optimized in the innovations filter since both frequency and phase components of the signal and coherent noise are utilized.

An example of the data obtained using the innovations filter is illustrated in Fig. 3. These data are of 22.8 ppmv chlorodifluoromethane in 100 kPa of argon and being excited with a pulsed TEA-CO_2 laser. With each pulse of the laser both the laser energy and the signal estimate are recorded. The scatter plots are of the same experiment but with different data-processing software. These data were not obtained simultaneously and therefore are of slightly different excitation energy ranges. The first plot is of signal estimates obtained with a pseudogated integrator. The gate was defined by summing a set range of channels in the digitized signal transient. The effective gate duration was 10 μs and was located at the signal maximum. (The entire transient can be seen in Fig. 4.) The second scatter plot utilized the matched filter. In this particular case there was no significant coherent interference and the innovations were made by orthogonalization of $s(t)$ to a constant valued base-line vector. Including this base line ensures that the filter will not respond to long-term base-line drift. In fact, even in the event of noticeable coherent interference, inclusion of this base-line offset function in the basis set results in better signal estimations using the innovations. The SNR of these data sets was calculated by using linear regression to a straight line and found from the squared average signal value to signal regression variance ratio. The SNR of the gated sampling data is 50 while that of the matched filter data is 6410. The SNR improvement of the matched filter over that of gated sampling is 128 for these data.

Innovations construction does not require the magnitude of $v_c(t)$ to be reproducible. The orthonormalization procedure is not dependent on the magnitude of the basis set functions. In cases where the magnitudes of $s(t)$ and $v_c(t)$

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitationnew.aip.org/termsconditions. Downloaded to IP: 129.123.125.251 On: Tue, 20 May 2014 17:40:18

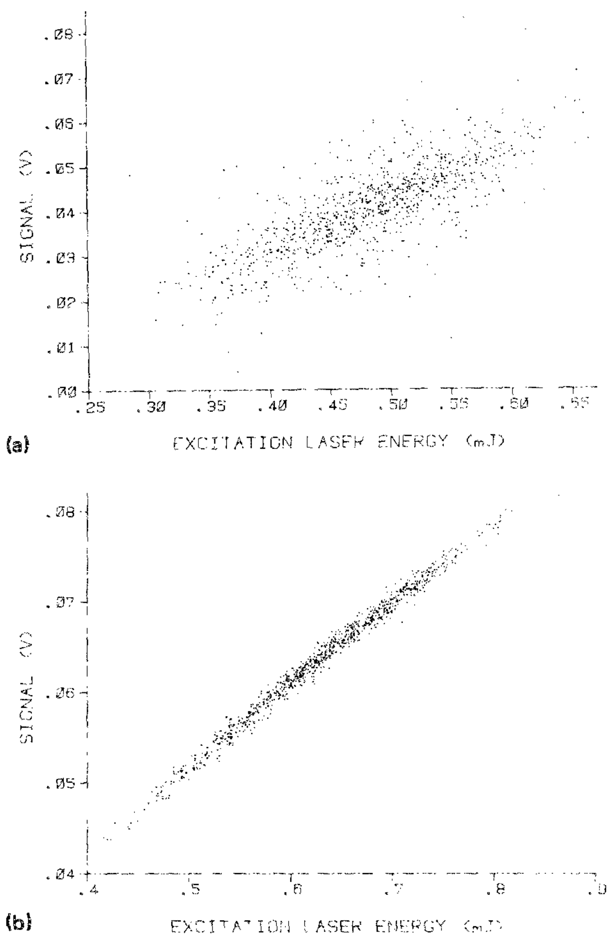

FIG. 3. Scatter plot indicating the photothermal signal-to-noise ratio for 22.8 ppm (v/v) of chlorodifluoromethane in argon at 100-kPa total pressure. The sample is excited at 1083.47 cm^{-1} using a TEA–CO_2 laser. The data of (a) used gated integration, while that illustrated in (b) used matched filtering for signal magnitude estimation.

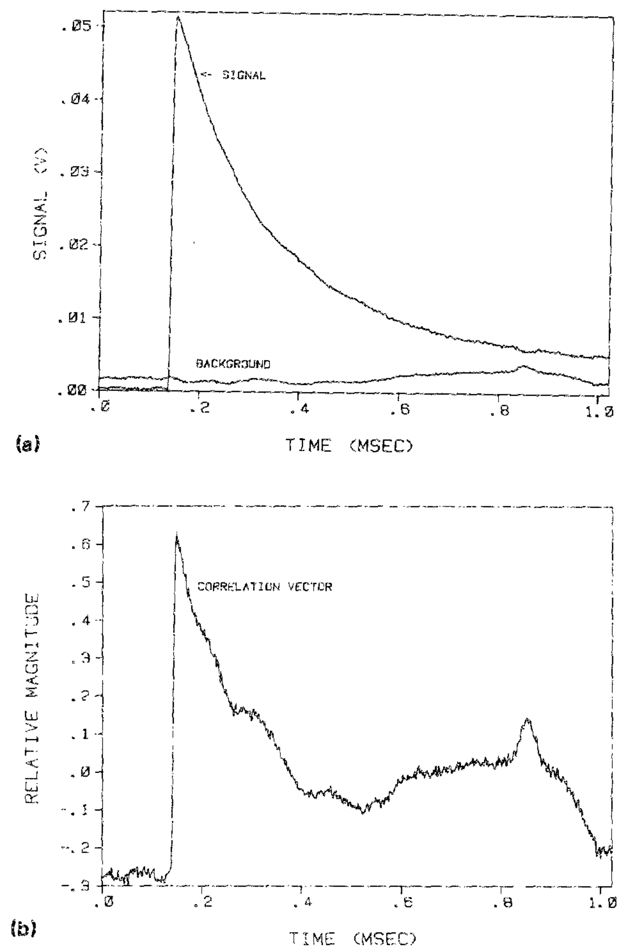

FIG. 4. An example of an empirically derived matched filter. Shown in (a) is the signal and the background obtained by multichannel averaging. Each averaged transient is a sum of 500 experimental cycles. The illustrated signal has been corrected for background interference. (b) shows the match filter correlation vector obtained by Gram–Schmidt orthogonalization. The apparent noise is due to the low signal-to-noise ratio in the averaged background.

vary substantially from cycle to cycle, the base-line subtraction technique described first may not work correctly since the accumulated estimate of $v_c(t)$ may not be the same as that found in the accumulated estimate of $x(t)$. Subsequent subtraction of $v_c(t)$ from $x(t)$ may not result in an accurate estimate of $s(t)$ in this instance. However, so long as the impulse-response function of $s(t)$ and $v_c(t)$ do not change with time, the innovations filter estimation of the signal magnitude will not vary with the magnitude of $v_c(t)$ since $i_s(t)$ is orthogonal to that component. Further, the accumulated estimate of $v_c(t)$ can be of arbitrary magnitude when used to construct the innovations filter. This relaxes the magnitude reproducibility criteria required for background subtraction method. The impulse-response of $s(t)$ must be known accurately for optimum magnitude estimation. One way to obtain an accurate estimation of $s(t)$ is to obtain the accumulated estimate under conditions where the SNR is very high and $v_c(t)$ is insignificant.

C. Some effects of coherent noise

The signal power obtained with the innovations filter must be less than or equal to that of the matched filter defined in Eq. (9) because of Schwartz' inequality [Eq. (8)]. The signal powers are equal only when there is no coherent interference in the expected input. In other words, the SNR obtained using the innovations filter will be less than that of the matched filter when no coherent interference is present, but was expected. In fact, using the signal innovations component defined by the Gram–Schmidt process,

$$i_s(t) = k\left[s_0(t) - \left(\int_0^\infty s_0(t)v_{c0}(t)dt\right)v_{c0}(t)\right], \qquad (23)$$

where $s_0(t)$ and $v_{c0}(t)$ are the square normalized expected signal and coherent impulse-response functions and k is the normalization constant, it is straightforward to show that the SNR ratio of the innovations to that of the matched filter is

$$\mathrm{SNR}_{\mathrm{inn}}/\mathrm{SNR}_{\mathrm{match}} = \left[1 - \left(\int_0^\infty s_0(t)v_{c0}(t)dt\right)^2\right]^2, \qquad (24)$$

when no coherent noise is present in the input. There is some consideration as to when to use the orthogonal filter method. If the coherent noise power is low relative to that of the

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitationnew.aip.org/termsconditions. Downloaded to IP: 130.102.135.251 On: Tue, 20 May 2014 17:40:18

signal, and the overlap integral between the expected signal and coherent noise is large, then the SNR of the resulting innovations filter may actually be less than that of the matched filter method which does not account for the coherent noise. However, both a decrease in SNR and an increase in signal estimate bias will occur if the matched filter is used and coherent interference is present. The bias is due to the finite overlap between the expected signal and coherent interference components of $x(t)$. Thus the signal estimate in Eq. (11) will be a sum of signal and coherent noise terms. If it is more important to reduce signal estimate bias, then the innovations filter should be used.

This is illustrated in the innovations filter construction shown in Fig. 4. Figure 4(a) shows the coherent background and signal vectors used to construct the innovation. As mentioned above, the signal was obtained by subtracting the signal averaged background $v_c(t)$ from the averaged signal plus coherent noise $x(t)$. In Fig. 4(b) the innovations filter vector is shown. The noise of the background is increased and manifested in this vector. This increase in noise is due to the orthogonalization process resulting in the innovations vector. One way to ensure that this type of noise amplification does not occur is to average the background data to the same SNR as that of the signal.

D. Advantage over gated sampling

The theoretical advantage to matched filter signal magnitude estimation over that of the sample and hold or gated integration methods can be formulated only for specific functions. A gated integrator of a sampling duration t_g sampling a system with stationary white noise statistics, will accumulate a total signal power proportional to the square of the integral of the signal and a noise power proportional to the gate time. The SNR enhancement for matched filter estimation over that of the gated integrator can be illustrated using the normalized exponential signal function, $s(t) = \exp(-t)$. For the matched filter, the corresponding signal and noise quantities are found by performing the integrations over the squared exponential function. The SNR of the matched filter estimation for a unit of signal magnitude is

$$\mathrm{SNR}_{\mathrm{match}} = (1 - e^{-2t_g})/2\sigma^2. \tag{25}$$

It is interesting to note that the SNR increases with increasing time. The longer the time of the filter, the better the SNR. This is interesting because the same is not true with the gated integration filter. In the latter, the SNR will decrease with increasing gate time. The SNR ratio of the matched filter to that of the gated integrator is found to be

$$\mathrm{SNR}_{\mathrm{match}}/\mathrm{SNR}_{\mathrm{gate}} = t_g(1 + e^{-t_g})/2(1 - e^{-t_g}). \tag{26}$$

In deriving this equation, the integration time was taken to be from zero to t_g for both filters. Although it is not easily seen, the SNR ratio steadily increases with time.[26] Further, this equation exhibits the intuitively correct behavior in the limit of an infinitely short sample time. For very short times, the SNR ratio is unity.

E. Considerations for discrete sampling devices

The transient digitizer is a discrete sampling device. The sampling function imposed on the signal by the digitizer may be characterized by three parameters: the sampling rate, the sampling aperture duration, and the total time sampled equal to the number of samples divided by the sampling rate. It has been stated that the SNR of the signal magnitude estimate does not change with the number of discrete recorder channels used to record a single transient.[24] The latter was found to be true only if the signal is band limited by a low-pass filter process prior to being recorded. If the low-pass filter frequency is the Nyquist frequency, then no increase in SNR can be obtained by sampling at faster rates. However, the limiting argument is one where a filter is utilized to pass only those frequency components that are synchronous with the experiment period. As a general rule, the error in this argument is that while reduction of the measurement bandwidth does reduce the sampled noise power, it can reduce the signal power as well. This error is illustrated particularly well in low duty cycle experiments where the signal power spectrum is primarily composed of high-frequency components which are greater than that of the experimental repetition frequency. In this instance, bandwidth-limited detection can all but eliminate the signal power since the majority of the signal power frequency distribution is beyond the cycle repetition frequency and this will result in a decrease in the SNR. Inspection of Eq. (26) in the limit as t_g approaches infinity will quickly dismiss this notion.

Consider an experiment where the total sampling time is fixed at t_s and the signal is sampled with N equally spaced gates or channels, with an interval of t_c. The gate duration t_g is assumed to be very small relative to t_c and thus can be considered to be a delta function. This is typical of transient digitizers using flash analog-to-digital converters of sample and hold circuits followed by a successive approximation converter. Also, after recording, the data are multiplied by the matched filter which, in this case, is assumed to be the expected signal. From the transform of Eq. (5), the noise power for stationary white noise using the expected signal at one phase or experimental time is

$$y_v^2 = \sigma^2 \sum_{n=0}^{N} s^2(nt_c), \tag{27}$$

the signal power at the same experimental time is from Eqs. (8) and (9),

$$y_s^2 = k\left(\sum_{n=0}^{N} s^2(nt_c)\right)^2, \tag{28}$$

and thus the SNR is,

$$\mathrm{SNR} = (k/\sigma^2) \sum_{n=0}^{N} s^2(nt_c). \tag{29}$$

It is not possible to estimate the SNR improvement as a function of N without knowledge of the expected signal. However, the ratio of two SNR at different N can be used to infer the predicted enhancement upon increasing N. Consider the case where N is doubled with a corresponding decrease in t_c,

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitationnew.aip.org/termsconditions. Downloaded to IP: 129.123.125.251 On: Tue, 20 May 2014 17:40:18

$$\mathrm{SNR}_{2N} = (k/\sigma^2) \sum_{n=0}^{N} \left\{ s^2(nt_c) + s^2\left(\left[n + \frac{1}{2}\right]t_c\right)\right\}. \quad (30)$$

Here, SNR_{2N} is the SNR with twice as many channels as in Eq. (29). All other definitions are the same. If N is large enough that the signal is well reproduced without aliasing, but still with a Nyquist frequency well below the instrumental frequency limit, then on the average,

$$s^2([n+1/2]t_c) = (1/2)\{s^2(nt_c) + s^2([n+1]t_c)\} \quad (31)$$

and the SNR improvement upon doubling N is

$$\frac{\mathrm{SNR}_{2N}}{\mathrm{SNR}} = \left\{2 \sum_{n=0}^{N} s^2(nt_c) - s^2(0) + s^2\left(\left[N + \frac{1}{2}\right]t_c\right)\right\} \Big/ \left(\sum_{n=0}^{N} s^2(nt_c)\right), \quad (32)$$

which for large N is approximately 2. Thus there is a SNR improvement in the power ratio on the order of N, corresponding to the usual square root law for SNR improvements in signal energy estimates.

Experimental verification that the SNR does increase as the number of data channels used per transient has been reported previously.[15] There is another improvement that will occur when more channels are used for signal recording. There is a constant quantization error in each digitized word due to the finite number of bits in the digital format. An improvement in the SNR will occur when several channels are used to record the same signal if the quantization error is random in the least-significant bit. This improvement will not be examined in this work. It is sufficient to point out this improvement and to recognize that the maximum number of channels should be used for digital matched filter processing such as the one described here.

IV. DISCUSSION

The time-domain matched filter is sensitive to both the frequency and phase of the signal being estimated and can be made independent of coherent noise interferences. This results in an enhancement of the signal estimate over filters operating in the frequency domain alone or those obtained with gated integration and sampled signal methods. This filter is not equivalent to frequency apodization of the transformed time-dependent data. Equivalent frequency-domain filtering can only be accomplished if the apodization function is time periodic. Further, because the filtering is performed with the time domain data, transformation to frequency domain is not required, and the signal magnitude estimate can be obtained rapidly in comparison to transform filter techniques.

The filter described above is not intended to be the final filter for a cyclostationary periodic process. Advanced adaptive filters such as the Kalman, adaptive Savitzky–Golay, and adaptive matched filters can be used after the optimal signal magnitude estimation of the real-time filter described here.[17–21] In fact, application of these postexperimental filters should be more valid when the signal estimate is obtained with this real-time filter since the innovations reduce, if not eliminate, in theory, measurement bias due to coherent noise. The resulting signal estimates should be characterized by stationary noise statistics from which the abovementioned advanced filters were derived.

Finally, it is interesting to note that this filter does not require a predictive estimate of the expected signal impulse-response. The phenomenological signal itself can be used to obtain the optimal estimate of the signal magnitude. The filter is thus adaptive in the sense that it may be changed on an experiment by experiment basis. This updating process requires little more effort than is normally required for multichannel averaging. However, unlike multichannel averaging, the advantage of this filter is that it can be used for optimal estimation of signals that vary in magnitude over time scales longer than the experimental cycle time and may be used to estimate correlations of nonlinear system excitation responses.

ACKNOWLEDGMENTS

Support for this work by the Office of the Vice President for Research, Utah State University and by CHE-8520040 awarded by the National Science Foundation is gratefully acknowledged.

[1] D. A. Skoog, *Principles of Instrumental Analysis*, 3rd ed. (Saunders College, New York, 1985).
[2] H. H. Willard, L. L. Merritt, J. A. Dean, and F. A. Settle, *Instrumental Methods of Analysis*, 6th ed. (Van Nostrand, New York, 1981).
[3] G. Kateman and F. W. Pijpers, *Quality Control in Analytical Chemistry* (Wiley, New York, 1981).
[4] G. M. Heiftje, Anal. Chem. **44**, 69A, 81A (1972).
[5] H. Doherty, Lasers Appl. **2**, 41 (1983).
[6] F. Castelli and R. D. Hefner, Rev. Sci. Instrum. **47**, 509 (1976).
[7] L. M. Jones, G. E. Leroi, C. A. Myerholtz, and C. G. Enke, Rev. Sci. Instrum. **55**, 204 (1984).
[8] D. Kumar, R. V. Nauman, and S. P. McGlynn, Rev. Sci. Instrum. **57**, 365 (1986).
[9] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 2nd ed. (McGraw-Hill, New York, 1984).
[10] G. R. Long and S. E. Bialkowski, Anal. Chem. **56**, 2806 (1984).
[11] G. R. Long and S. E. Bialkowski, Anal. Chem. **57**, 1079 (1985).
[12] G. R. Long and S. E. Bialkowski, Anal. Chem. **58**, 80 (1986).
[13] S. L. Nickolaisen and S. E. Bialkowski, Anal. Chem. **57**, 758 (1985).
[14] S. L. Nickolaisen and S. E. Bialkowski, Anal. Chem. **58**, 215 (1986).
[15] S. L. Nickolaisen and S. E. Bialkowski, J. Chem. Info. Comp. Sci. **26**, 57 (1986).
[16] S. E. Bialkowski, Anal. Chem. **58**, 1706 (1986).
[17] M. U. Bromba and H. Ziegler, Anal. Chem. **55**, 1299 (1983).
[18] M. U. Bromba and H. Ziegler, Anal. Chem. **55**, 648 (1983).
[19] H. N. Poulisse, Anal. Chem. Acta **112**, 361 (1979).
[20] S. C. Rutan and S. D. Brown, Anal. Chem. Acta **160**, 99 (1984).
[21] S. A. Dyer and D. S. Hardin, Appl. Spectrosc. **39**, 655 (1985).
[22] A. D. Poularikas and S. Seely, *Signals and Systems* (PWS, Boston, 1985).
[23] G. H. Golub and C. F. Van Loan, *Matrix Computations* (John Hopkins University, Baltimore, MD, 1983).
[24] R. R. Ernst, Rev. Sci. Instrum. **36**, 1689 (1965).
[25] P. R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences* (McGraw-Hill, New York, 1969).
[26] One reviewer correctly points out that the infinite-time matched filter SNR is only 8.83% better than the optimum finite-time gated integration SNR. The extra 8.83% gained is probably not worth the infinite wait!

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitationnew.aip.org/termsconditions. Downloaded to IP: 129.123.125.251 On: Tue, 20 May 2014 17:40:18