

Real-time signal-processing concepts for trace-gas analysis by diode-laser spectroscopy

Peter Werle, MEMBER SPIE

Bodo Scheumann

Josef Schandl

Fraunhofer Institut für Atmosphärische

Umweltforschung

Kreuzeckbahnstr. 19

82467 Garmisch-Partenkirchen, Germany

E-mail: werle@ifu.fhg.de

Abstract. Tunable-diode-laser absorption spectroscopy fulfills the major requirements for trace-gas analysis: sensitivity, specificity, high detection speed, and the possibility of simultaneous *in situ* measurements. The well-known limitations for low-concentration measurements become more and more dominant at sub-part-per-billion levels, where sensitive spectrometers are often influenced by noise, drift effects, and changes in the spectral background structure. While many improvements in instrument development focus on optimizing electronics and optical components, much less effort has been put into postdetection signal processing and adaptive control. Therefore, a transputer-based platform has been developed which allows control of most relevant parameters of a diode-laser spectrometer. Fluctuations in the signal amplitude as well as drift and jitter effects in the frequency domain can cause a significant degradation of system performance and therefore determine the ultimate detection limit. A signal-processing concept with novel aspects for tunable-diode-laser spectroscopy is presented and discussed.

Subject terms: diode laser spectroscopy; trace gas analysis; adaptive signal processing.

Optical Engineering 33(9), 3093–3105 (September 1994).

1 Introduction

The measurement of atmospheric trace gases imposes high demands on instrumentation. Fast, accurate, and rugged instruments are needed for airborne applications and trace-gas flux measurements. Ultrasensitive instruments are required to measure free radicals and reactive species in the atmosphere, where the measurements should be free of interference from other atmospheric constituents. The great number of gaseous pollutants and their generally low, variable concentrations with large local differences pose challenging requirements for analytical techniques.¹ Thus, sensitive, selective, and mobile or even portable instruments with a large dynamic range are needed. Modern research in atmospheric chemistry requires highly sensitive techniques for the measurement of concentrations of free radicals, which determine the rate of photochemical destruction of most atmospheric pollutants.

Tunable-diode-laser absorption spectroscopy (TDLAS) is now frequently used for the measurement of trace-gas pollutants in the atmosphere. These spectrometers use multipass absorption cells to achieve high sensitivity.² To alleviate problems of absorption-line overlap, these absorption cells are usually operated at low pressure, where the linewidth is Doppler limited. In most sensitive instruments the diode laser is repetitively tuned across an absorption line of a target molecule and the absorption spectra are averaged over a specified time interval. Additional modulation techniques are used

to reduce the $1/f$ laser noise. With wavelength modulation, detection limits of the order of, typically, 0.1 ppbv (1 ppbv = 10^{-9} volume mixing ratio) have been achieved for many smaller molecules in the air with spectrum averaging time of a few minutes. Although these detection limits are sufficient for many applications, still better detection limits are required by modern atmospheric research for the determination of concentrations of free radicals, which play a decisive role in the destruction of almost all atmospheric pollutants and in the formation of atmospheric ozone. The required detection limits are of the order of a few pptv for HO₂ and even lower for HO radicals.³ Substantial improvements in TDLAS detection limits have been obtained by introducing high-frequency modulation techniques.^{4,5} A potential sensitivity improvement of up to two orders of magnitude in comparison with conventional derivative ($2f$) spectroscopy has been derived from wideband noise characteristics of lead-salt diode lasers. This sensitivity improvement can be achieved by increasing the modulation frequency from the $1/f$ -noise-dominated region (10 kHz) into a shot-noise-limited domain (about 100 MHz). In terms of optical density, the best detection limits obtained so far with a high-frequency modulated TDLAS instrument with a multiple-pass absorption cell were of the order of 10^{-6} with a detection bandwidth^{6,7} of 1 Hz. From a theoretical point of view, the signal from a perfectly stable system could be averaged infinitely. Infinite averaging should lead to extremely sensitive measurements if limitations on the dynamic range of the system can be neglected. Unfortunately, real systems are not stable indefinitely, and a fundamental question is how

Paper 04024 received Feb. 1, 1994; accepted for publication Mar. 24, 1994.
© 1994 Society of Photo-Optical Instrumentation Engineers. 0091-3286/94/\$6.00.

long signals can be averaged to achieve an optimum sensitivity with a given spectrometer. It is obvious that every real unstable system will have an optimum averaging time determined by the drifts in the system, such as temperature drifts, moving fringe structures, and other background changes.⁷ To achieve the best possible sensitivity, the instrument should be made as stable as possible and the measuring time should be as short as possible. While many improvements in TDLAS focus on optimizing electronics and optical components, much less effort has been put into post-detection signal processing.^{8,9} It is the purpose of this paper to discuss signal-processing strategies for tunable-diode laser spectroscopy and real-time signal analysis applications using sophisticated data-processing techniques.

Trace-gas measurements near to the detection limit are usually performed by measuring alternately the *ambient air spectrum*, and the spectrum of zero air, i.e., air devoid of the target substance, which we will refer to as the *background spectrum*. A prerequisite for quantitative measurements is a *calibration spectrum*, which for example can be obtained from a permeation-based calibration system. From time to time such a spectrum is recorded and stored in the computer memory for signal processing. Finally, a *reference spectrum* is used for online monitoring of the laser amplitude and frequency fluctuations. Therefore a fraction of the total laser intensity is coupled out of the laser beam by a beam splitter, and after it passes through a reference cell filled with a high concentration of the trace gas under investigation, a strong signal with a high signal-to-noise ratio is generated. As noise, spikes, interferences, and drifts superimposed on the desired laser current manifest themselves as variations in the laser frequency, it is important to monitor these parameters. Typical tuning rates of lead-salt semiconductors used in spectroscopic instrumentation² vary from several hundred megahertz up to 2 GHz/mA. Therefore, a fluctuation in the laser current of only a few microamperes can cause frequency fluctuations of several megahertz.

Due to the high modulation frequencies (megahertz) and fast demodulation in FM spectrometers, the laser can be tuned across the spectral feature of interest with frequencies in the kilohertz range with these instruments. We preferred high scanning frequencies to discriminate against low-frequency drifts and vibrations. Scanning frequencies of the order of kilohertz, however, result in a data acquisition rate of the order of several hundred kilohertz. To cope with the high data acquisition rate, the spectral data are acquired and averaged by a transputer, and the real-time signal analysis is made by an additional processor. The instrument automatically provides signal, background, and calibration spectra consisting of up to 4096 channels and allows online averaging of more than 10^4 individual spectra. The integrated laser control system is based on flexible hardware using parallel data processing for each channel.

System stability has been investigated by monitoring several parameters during measurements with the FM TDLAS system.⁷ The analysis of such measurements reveals a good correlation between the deviation from the line center and the confidence range obtained from a fit. The confidence range under "locked" conditions was less than 50 parts per trillion (ppt), and it may degrade by up to a factor of 10 when the deviation is large.⁷ This finding illustrates the importance of a highly stable, transient-free laser current source and calls

for improved line-locking schemes and advanced post-detection signal processing.

At low ambient trace-gas levels, the demodulated signal usually has no clean structure and a correct calculation of the gas concentration is impossible. The signal-to-noise ratio (SNR) is a convenient way to describe the "cleanliness" of a given signal level. It is simply the signal voltage (or power) divided by the rms noise voltage (or power). A spectrum obtained with a SNR of 100 or more has a very clean pattern, with negligible noise. At a SNR of 10 the pattern is still very clear, but a little fuzzy. At a ratio of 3 the pattern is bad, and at a ratio of 1 the signal is nearly lost. A set of typical spectra is shown in Fig. 1. The dotted traces show a downscaled calibration spectrum. The raw spectra show no fringe structure, but a positive and a negative slope are shown superimposed on the upper two traces. This slope can be interpreted as part of a fringe with a broad free spectral range, which generates a linear background structure on this scale. Background changes can, at least in part, be attributed to fringe movements. In many cases a more or less pronounced curvature has been observed. Such varying slopes can also be generated from pressure broadened atmospheric absorption lines. Such an uncertain situation, where an *a priori* knowledge of the spectra and its background is necessary, calls for the introduction of "adaptivity." A need for self-adaptive algorithms having some learning capabilities arises in the control of such processes, which are time varying, are non-linear, and have unknown dynamics with unknown disturbances acting on them. For such a complex problem, no analytical solution can be found. Although a potential *a priori* control structure and some performance criteria can be defined, it is generally not possible to specify, in advance, the parameters within this structure.

The general structure of a signal-processing unit for diode-laser spectroscopy is shown in Fig. 2. It can be divided into three main elements. The first element is a *signal averager*, which is required for SNR improvement and online correction of fast frequency fluctuations in the ambient spectra. The smoothed output of the signal averager is fed into an *adaptive digital filter*, where either a signal-synthesis adaptive technique (SSAT) or a parameter adaptive technique (PAT) can be applied. The fundamental difference between these approaches is that the first results in the generation of a signal, whereas the second is based on the alteration of controller parameters. In general both mechanisms yield the same system, and mostly they are based on the concept of minimizing a performance feedback criterion. Finally, a normalization process is required to correct for amplitude variations and to determine the concentration values from ambient spectra. In the next sections, novel signal-processing aspects suited to TDLAS will be discussed.

2 A Digital Signal Averager with On-Line Drift Correction

It has been mentioned in the previous section that ambient signals at low concentration levels have a poor SNR, and therefore signal averaging is necessary prior to further signal processing. Input parameters for such an averager are the number S of spectra that have to be block averaged, the number N of channels per spectrum, and the data x_i from the analog-to-digital converter (ADC). As output the averager calculates the mean \bar{x} and the variance σ^2 of the input time-

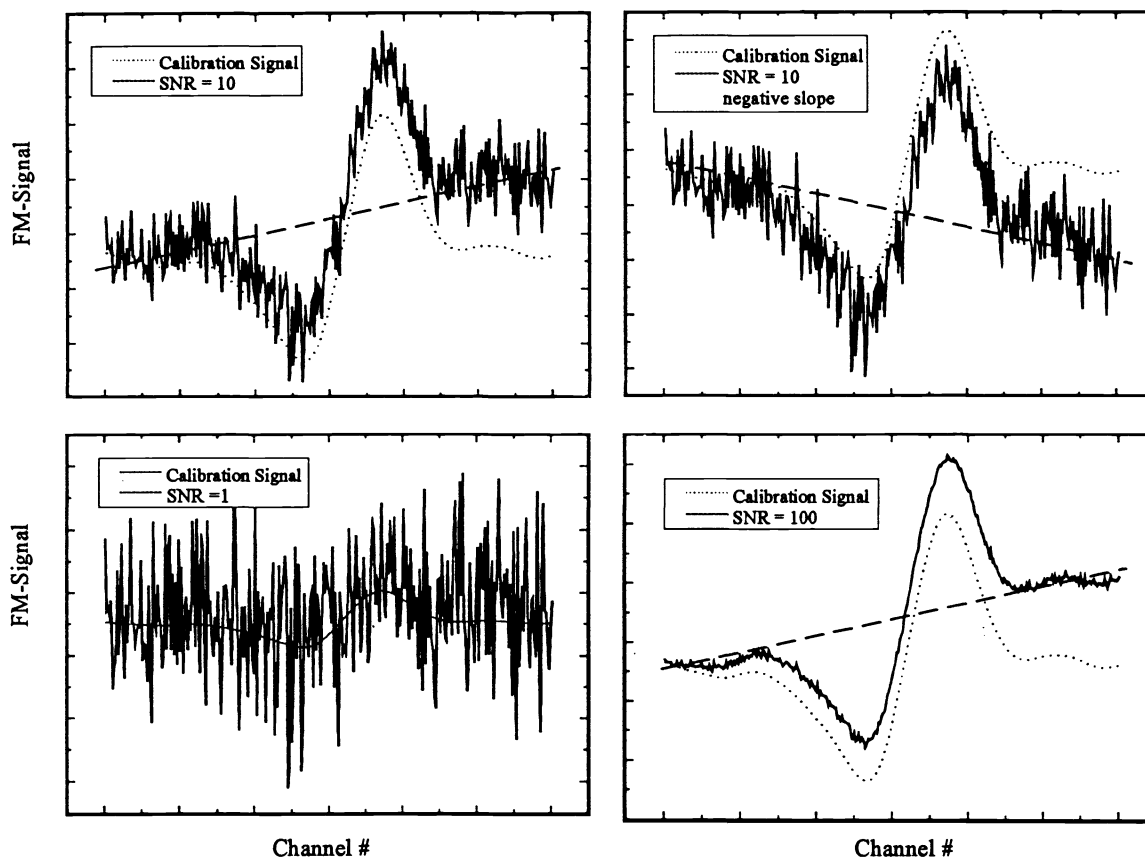


Fig. 1 Typical FM signal at different SNRs with a superposition of a positive and a negative linear slope.

series data. The duty cycle has a significant influence on the system performance,⁷ and, within certain limits, this performance does not depend on the total data acquisition rate. Therefore this averaging process runs continuously, parallel to the subsequent analysis, and dead-time effects are minimized, i.e., while the averaging of a data block is running, the averaged spectrum from the previous block is analyzed.

Another important requirement is that all signals should have approximately the same SNR. Therefore, the averager performs the same process with all input data, regardless of their nature. Digital averaging reduces the electronic system bandwidth Δf_{sys} within certain limits,⁷ according to

$$\Delta f_{\text{eff}} = \frac{\Delta f_{\text{sys}}}{S},$$

where S is the number of averaged spectra. As background and calibration spectra are required for the signal processing, they are temporarily stored in the computer memory, where they are updated at certain time intervals. These time intervals are determined by measuring the system stability, using calculations of the Allan variance to derive the optimum integration time and the resulting interval for ambient and background measurements.⁷

It has already been mentioned that variations in the laser current due to noise, spikes, drifts, 50- or 60-Hz hum, and other interfering signals have a dramatic influence on the system performance, especially on the confidence range of

the measurement and therefore on the detection limit of the system.⁷ An order of magnitude in performance can easily be gained or lost, depending on the stability of the laser. One approach to these problems is the improvement of the laser current supply (e.g., by reducing the output bandwidth), but practical limits are set by the scan frequency of the system. A different approach, which we propose, is to correct observed drifts on line during the averaging process.

The principle function of the *online shifter*⁹ with a single-tone FM signal is illustrated in Fig. 3. A fixed trigger initiates each scan. Discrete frequency interferences superimposed on the laser current cause a signal drift or jitter along the x axis, which is synchronous in the sample and in the reference channel. Coaveraging of these spectra leads to a degraded system performance due to the broadening of the averaged spectrum and an offset, indicated by the two arrows in Fig. 3. The application of the on-line shift procedure eliminates or at least corrects the disturbances introduced by interfering signals. Therefore, a latch pulse is generated in the reference channel, which has to have a high SNR for stable trigger generation. This is easy to obtain by using a reference absorption cell with a high concentration of the trace gas under investigation. Discrete frequency interferences in the laser current supply and even slow drifts can be suppressed, and an active line-locking mechanism only has to prevent the signal from drifting out of the monitoring window. While the averaged spectrum should be centered on the dashed line, discrete frequency interference or spikes generate a shift in

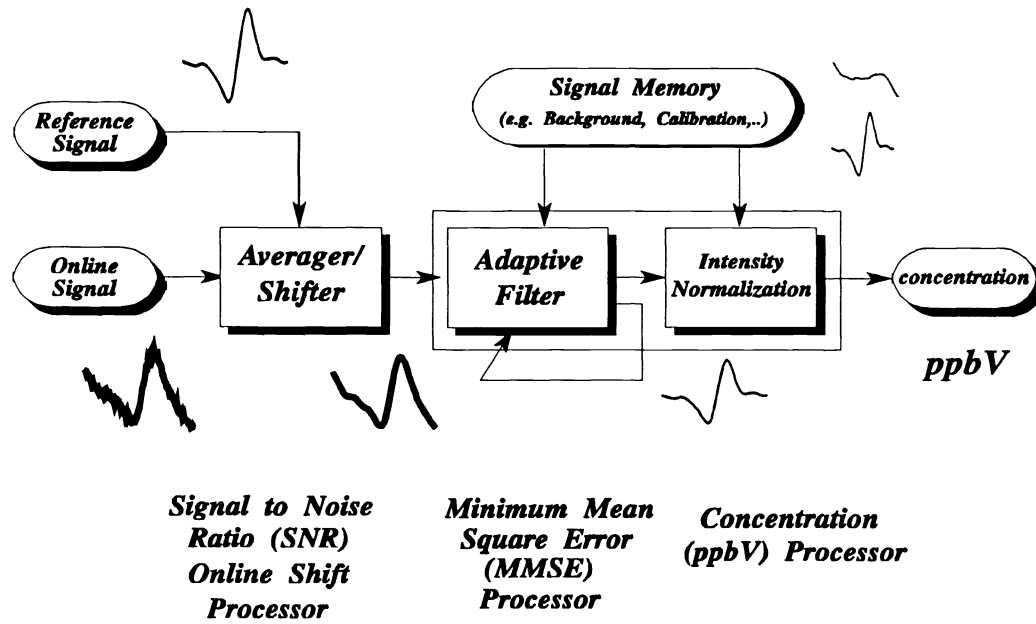


Fig. 2 A general signal-processing concept for tunable-diode-laser spectroscopy that allows correction of variations in the frequency scale (on-line shifter) and in the signal amplitude (normalization).

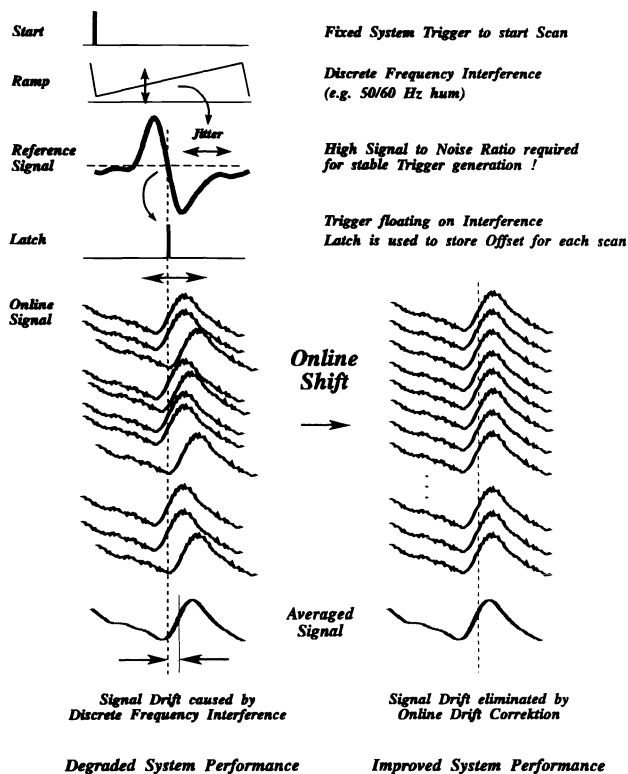


Fig. 3 Principle operation of the on-line shifter mechanism.

the averaged spectrum, which degrades the system performance; this will be discussed in the next section.

A general implementation of the on-line shift process is illustrated in Fig. 4. To minimize dead-time effects⁷ a double buffer system should be applied. While a currently digitized spectrum is stored in buffer 1, buffer 2 can be added in parallel

to the average memory. To determine the amount of shift required for the online correction, all registers and variables have to be set to zero. A register has to be loaded with a certain channel (say 64). If we assume 128 channels for the spectrum, the preset value is at the center of the spectrum. The start pulse enables an address counter to increment with each clock pulse. The address counter points to the current buffer address where the digitized on-line signals are stored. With the occurrence of the latch signal, which can be derived from the reference signal, the current buffer address is stored in the index register (say 65). In the first scan nothing happens in the parallel process operating on buffer 2, because only operations of the type $0 + 0$ will be carried out. The next cycle is initiated again by a start impulse, which simultaneously changes the sign of the contents of the index register (65) and shifts into the index' register. Summing of the index' register (-65) and the preset register (64) generates the jitter-induced offset (-1). As the buffers are swapped with the start signal, the on-line data will now be written into buffer 2. While the address counter points to the current buffer address (CBA), the current memory address (CMA) for the averaging process can be calculated by adding an offset (-1) and CBA, and the shift (in this example 1 channel) can be corrected. It is obvious that the quality of the shift mechanism depends on the number of channels in the ambient spectrum. With 100 channels a shift resolution of 1% is available. For a given number S of averages, $S + 1$ buffers have to be processed, due to the use of a dual buffer system. For a large number of averages the dead time can be neglected and the averaged signal is available in real time.

Figure 5(a) shows some experimental data, where the implemented shifter acts on an on-line two-tone FM signal. Without the on-line shifter a significant jitter can be observed when several spectra are plotted on the same graph, while with the activated shifter, drift or jitter reduction is significant. This can be seen more clearly in Fig. 5(b), where calculated

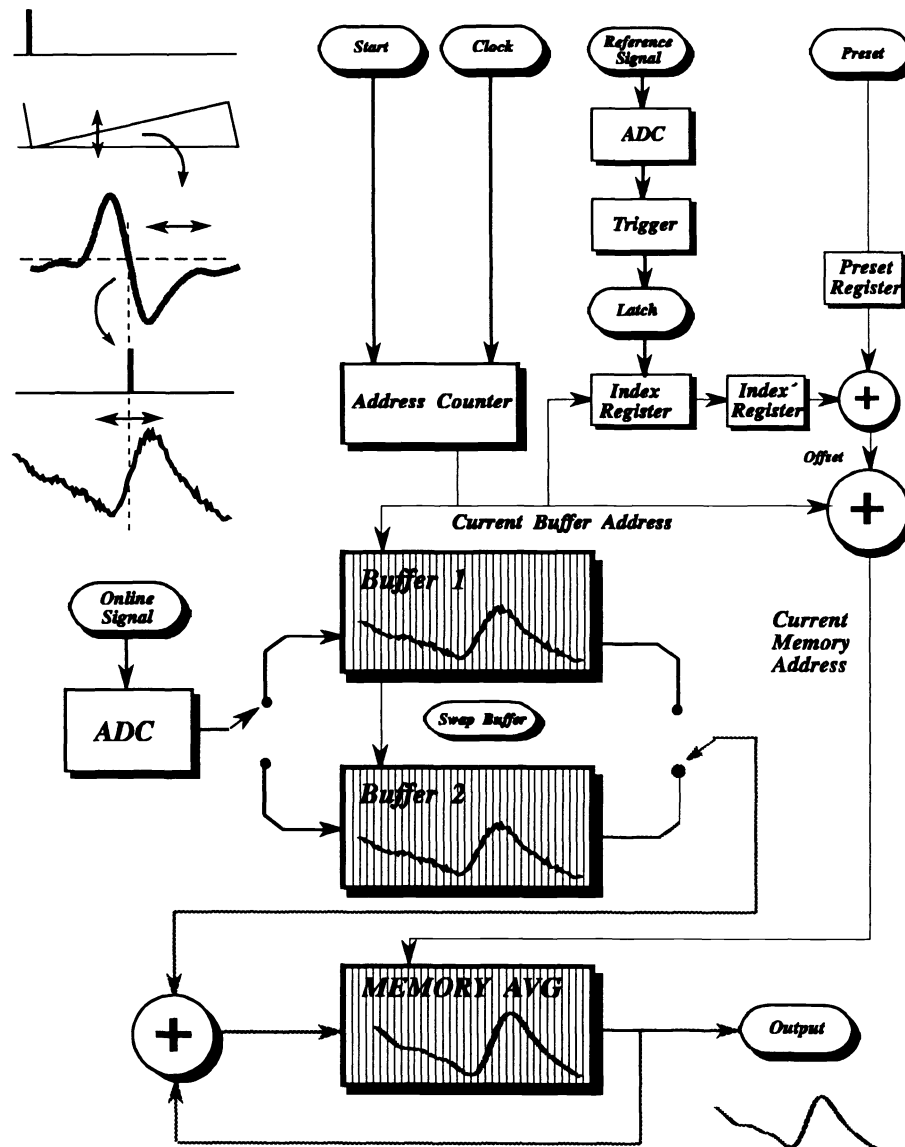


Fig. 4 On-line shifter implementation with a double buffer for optimum duty cycle.

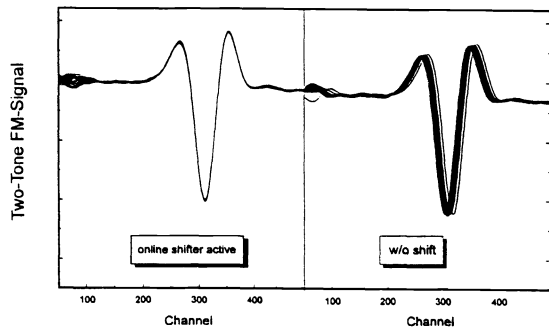
concentration values from a calibration source are plotted as a function of time, with the shifter on and off. When the shifter is active, a stable signal can be observed, and the confidence range shown in the lower trace is at a reasonably low level and does not change with time. When the on-line shifter is turned off, the calculated concentration gives significantly lower values, and much higher associated errors are seen in the lower trace. It should be noted that the true concentration value of 1 may not be within the confidence range of the calculated concentration value. These data demonstrate the significance of the correlation between signal stability in frequency space and system performance, and this calls for increased stability in the laser power supply and advanced signal-processing strategies.

3 Adaptive Signal Processing Techniques

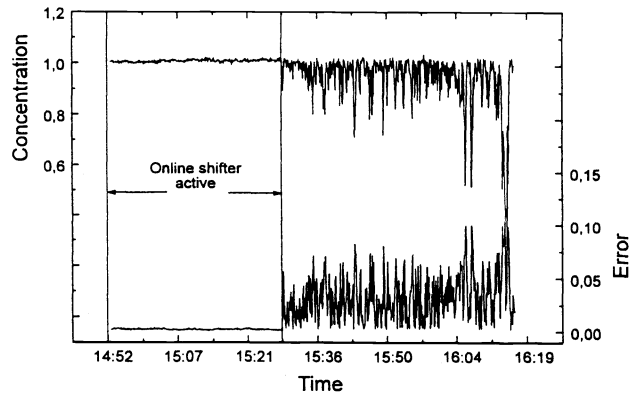
The essential and principal property of an adaptive system is its time-varying, self-adjusting performance. The need for such performance may readily be seen by realizing that if the

designer develops a system with a fixed design that he considers optimal, the implications are that the designer has foreseen all possible input conditions and knows what the system should do under each of those conditions. The designer has then chosen a specific criterion whereby performance is to be judged, such as the amount of error between the output to the actual system and that of some selected model or "ideal" system. Finally, the designer has chosen the system that appears best according to the performance criterion selected. In many systems, the complete range of input conditions may not be known exactly, or the conditions may change from time to time. In such circumstances, an adaptive system that continually seeks the optimum within an allowed class of possibilities (e.g., filters) would give better performance than a system of fixed design.

The adaptive linear combiner, or nonrecursive adaptive filter, is fundamental to adaptive signal processing.¹⁰ It appears, in one form or another, in most adaptive filters and systems, and is the single most important element in "learn-



(a)



(b)

Fig. 5 (a) On-line shifter operation on two-tone FM spectra. (b) System performance with and without the on-line shifter.

ing” systems and adaptive processes in general. Because of its nonrecursive structure, the adaptive linear combiner is relatively easy to understand and analyze. In essence it is a time-varying, nonrecursive digital filter where, for specific applications, criteria exist to define the “best” performance. A diagram of the general form of the adaptive linear combiner is shown in Fig. 6. The input signal vector $x_{1n}, x_{2n}, \dots, x_{Mn}$, a corresponding set of adjustable weights $w_{1n}, w_{2n}, \dots, w_{Mn}$, a summing unit, and a single output signal y_n are defined, where n is the time index. A procedure for adjusting or adapting the weights is called a “weight adjustment” or “adaptation” procedure. The combiner is called “linear” because, for a fixed setting of the weights, the output is a linear combination of the input components. However, when the weights are in the process of being adjusted, they too are a function of the input components; then the output of the combiner is no longer a linear function of the input. There are two important ways to interpret physically the elements of the input vector. First they may be considered to be simultaneous inputs from M different signal sources (multiple input) or they may be interpreted as M sequential samples from the same signal source (single input). In the single-input case the adaptive processor can be implemented with an adaptive linear combiner and unit delay elements, and is called an adaptive transversal filter. In some multiple-input systems, a bias weight is needed that simply adds a variable bias (e.g., a dc offset) to the sum y_n . We obtain this easily by setting the first input element x_{1n} permanently to 1. For a multiple input signal we obtain the input-output relationship

$$y_n = \sum_{v=1}^M w_{vn} x_{vn} = w_n^T X_n \quad (\text{multiple input}) .$$

For a single input system, the relationship is

$$y_n = \sum_{v=1}^M w_{vn} x_{n-v} = w_n * x_n \quad (\text{single input}) ,$$

where an asterisk denotes convolution and x_{n-v} represents a replica of the signal x_n shifted to the left along the time axis by the amount v . With an input vector $X_n = [x_{1n}, x_{2n}, \dots, x_{Mn}]^T$ and a weight vector $W_n = [w_{1n}, w_{2n}, \dots, w_{Mn}]^T$ we can write $y_n = W_n^T X_n$, where T stands for “trans-

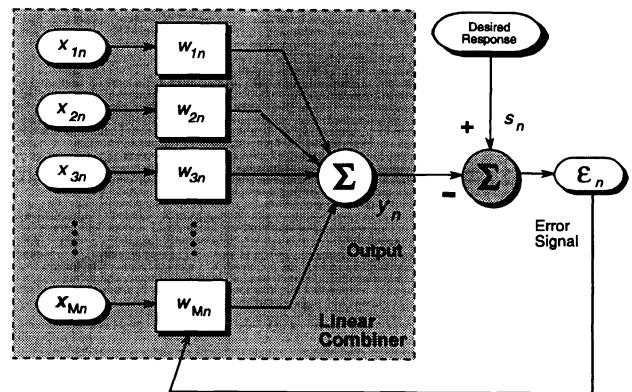


Fig. 6 General form of a multiple-input adaptive linear combiner with desired response and error signal.

pose”; X_n is actually a column vector. With closed-loop or performance feedback systems the weight vector depends on the output signal as well as on other data. Generally, the other data for the adaptive linear combiner include a “desired response” or a “training signal.” In the adaptation process with performance feedback, the weight vector of the linear combiner is adjusted to cause the output y_n to agree as closely as possible with the desired response signal. This is accomplished by comparing the output with the desired response to obtain an “error” signal and then adjusting or optimizing the weight vector to minimize this signal. In most practical instances the adaptive process is oriented toward minimizing the mean square value or average power of the error signal.

Although several ways of defining a performance index have evolved, the quadratic performance criterion has been chosen as the most satisfactory because of its attractive mathematical properties. We can write for the error signal

$$\epsilon_n = s_n - y_n$$

or

$$\epsilon_n = s_n - W_n^T X_n .$$

To develop an adaptive algorithm it is necessary to minimize

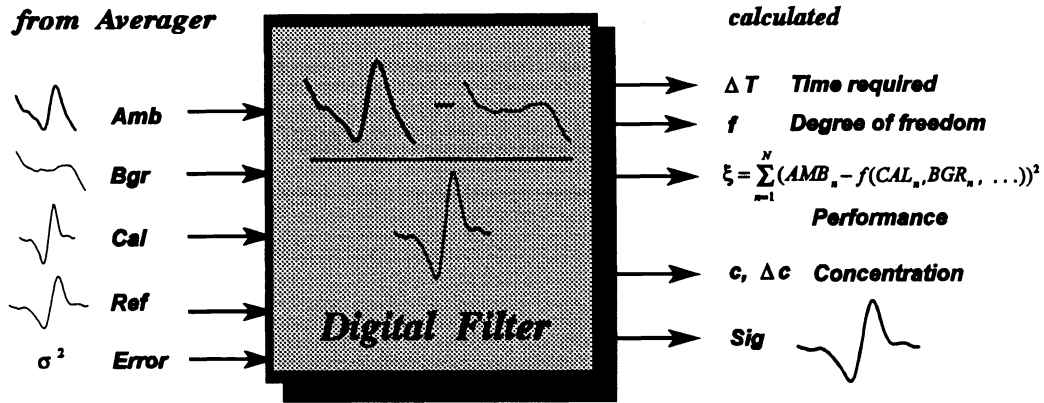


Fig. 7 Block diagram of a digital filter with input and output parameters.

the mean square error (MSE) ξ , which is the expectation value $E[\]$ of the squared error signal ϵ_n^2 :

$$MSE \equiv \xi = E[\epsilon_n^2] = \sum_{n=1}^N (s_n - W_n^T X_n)^2$$

It turns out that the performance surface of the MSE is an upward concave hyperparaboloid. In the case of two weights, the point at the bottom of the paraboloid gives the optimal weight vector W^* and the minimum mean square error (MMSE). Many useful adaptive processes that cause the weight vector to seek the minimum in the performance surface do so by gradient methods. If we use the gradient (∇) to find the MMSE by calculating the derivative

$$\nabla_{\mu} \xi \equiv \frac{\partial \xi}{\partial w_{\mu}} = 0$$

and solving the M resulting simultaneous equations, we can adjust the elements of the weight vector for optimum performance.

We now turn to the description of some digital filters that can be used for spectroscopic applications. The input vectors that are available are the reference signal (Ref), the background signal (Bgr), a background corrected calibration signal (Cal), and the actual ambient signal (Amb). Additionally, a vector containing information about the variances σ^2 of the averaged signals may be available. Such a general adaptive filter is shown in Fig. 7. The output parameters are the filtered signal (Sig), which serves as a "desired signal," the calculated concentration value c , and the corresponding error Δc , as well as performance criteria such as ξ and the time ΔT required for calculation. The desired signal S should contain a fraction of the calibration signal, the background, and some possible other contributions, which have to be specified in detail.

Therefore, in general, we can write $S_n = f(Bgr_n, Cal_n, \dots)$, and for the performance criterion

$$\xi = \sum_{n=1}^N (Amb_n - S_n)^2$$

An adaptive digital filter should be able to "learn" to synthesize the optimum signal by using a performance criterion

and a strategy to minimize the cost function. The function f , for example, can be a linear combination of the arguments; then $S_n = Bgr_n + c \cdot Cal_n$. This frequently used approach for adaptive signal processing is the *linear regression scheme*, which is a least-mean-square (LMS) algorithm (Fig. 8). This is the conventional approach in diode laser spectroscopy, where a previously taken background spectrum (Bgr) is subtracted from the actual ambient spectrum (Amb) and then fitted to a calibration spectrum. Formally, we have to minimize the performance function

$$\xi = \sum_{n=1}^N [(Amb_n - Bgr_n) - c \cdot Cal_n]^2 \rightarrow \text{Min}$$

where $S_n = Bgr_n + c \cdot Cal_n$ is the desired signal. To determine the concentration we have to find the minimum of $\xi(c)$, i.e., set the partial derivative of ξ equal to zero:

$$\frac{\partial \xi}{\partial c} = 2 \sum_{n=1}^N \{-Cal_n \cdot Amb_n + Cal_n \cdot Bgr_n + c \cdot Cal_n^2\} = 0$$

$$\gamma_{CA} = \sum_{n=1}^N Cal_n \cdot Amb_n$$

$$\gamma_{CB} = \sum_{n=1}^N Cal_n \cdot Bgr_n$$

$$\gamma_{CC} = \sum_{n=1}^N Cal_n^2$$

The calibration factor $c = (\gamma_{CA} - \gamma_{CB}) / \gamma_{CC}$ is the ratio of the background-corrected ambient signal to the calibration signal, and therefore reflects the concentration of the gas. This algorithm is insufficient, because it is sensitive to any slope and curvature in the background structure and extremely sensitive to even small drifts with time, especially between calibration and ambient spectra. Depending on the direction of the slope, the signal is either under- or overestimated. It even can happen that the true value 0.5 does not lie within the confidence range reported from the fit, as indicated in Fig. 9(a), where results of a linear regression routine are plotted as a function of the slope added to the spectrometer signal

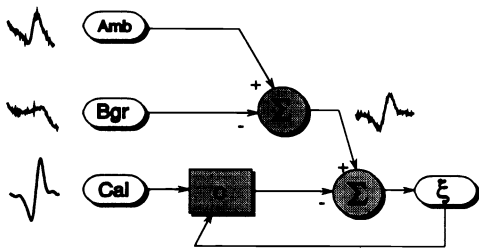
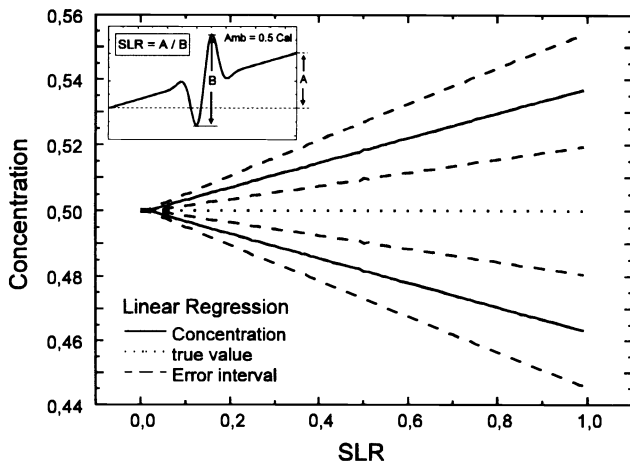
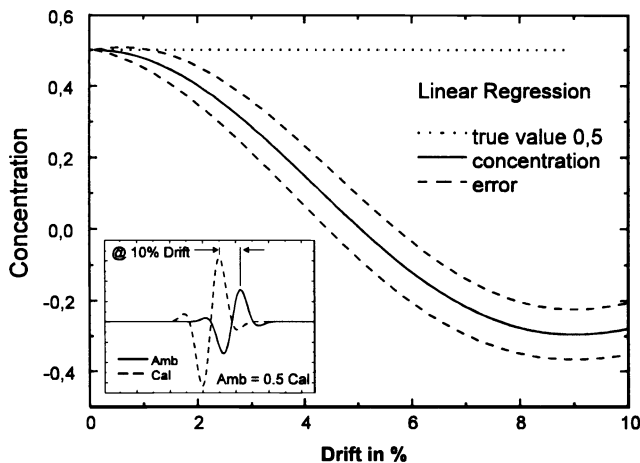


Fig. 8 A widely used linear regression scheme drawn as a basic adaptive filter.



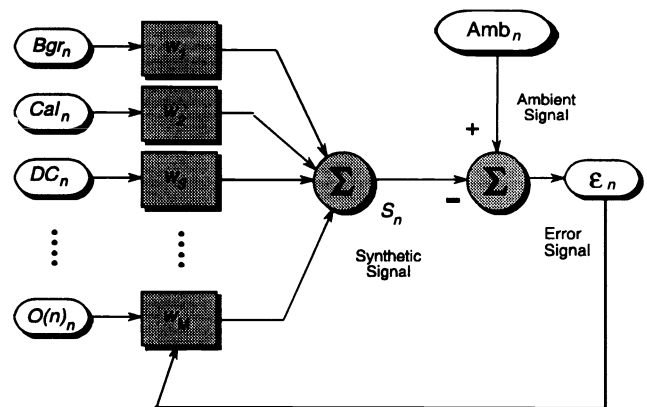
(a)



(b)

Fig. 9 (a) Calculated "concentration" as a function of the signal-to-linear-ramp (SLR) ratio. (b) Calculated "concentration" as a function of the drift between ambient and calibration signal.

with a high SNR. The dependence of the calculated concentration as a function of the drift is plotted for drifts up to 10% in Fig. 9(b). In both figures the ambient signal was assumed to be 50% of the calibration signal, and therefore the "true" concentration is 0.5. Both effects can, under worst-case conditions, lead to completely wrong concentration values. A remarkable result of this analysis is that the errors reported from the fit algorithm remain constant when



(a)

$$\begin{pmatrix} Amb_1 \\ Amb_2 \\ \dots \\ Amb_N \end{pmatrix} = \begin{pmatrix} Bgr_1 \\ Bgr_2 \\ \dots \\ Bgr_N \end{pmatrix} \begin{pmatrix} Cal_1 \\ Cal_2 \\ \dots \\ Cal_N \end{pmatrix} \begin{pmatrix} DC \\ DC \\ \dots \\ DC \end{pmatrix} \dots \begin{pmatrix} w_1 \\ w_2 \\ \dots \\ w_M \end{pmatrix}$$

(b)

Fig. 10 (a) Multiple linear regression scheme based on a linear combiner implemented for the signal synthesis adaptive technique. (b) Illustration of matrix elements for MLR scheme.

drift values are above 2 to 3%. Therefore, a system operator or control software cannot identify bad data due to an increasing error.

A promising approach is the use of a multiple linear regression (MLR) scheme.^{11,12} This is an extension of the linear regression scheme and offers more degrees of freedom for the fit process [Fig. 10(a)]. While the linear regression discussed previously uses only information about the background and the structure of the calibration signal, the multiple linear regression scheme allows more parameters to be added. A typical situation in diode-laser spectroscopy is one where a more or less strong fringe structure is superimposed on the ambient signal. If the free spectral range of the fringe is greater than the ambient signal, the fringe can manifest itself as a slope or another background structure with a curvature. A slowly moving fringe structure therefore generates a time-varying background structure, which calls for adaptive signal processing in the sense previously discussed, and we can write for the desired signal

$$S_n = w_1 \cdot Bgr_n + w_2 \cdot Cal_n + w_3 \cdot DC_n$$

$$+ w_4 \cdot Lin_n + \dots + w_M \cdot O(n)_n,$$

and for the performance function, which has to be minimized,

$$\xi = \sum_{n=1}^N \{ Amb_n - [w_1 \cdot Bgr_n + w_2 \cdot Cal_n + w_3 \cdot DC_n + w_4 \cdot Lin_n + \dots + w_M \cdot O(n)_n] \}^2,$$

where DC can be set to unity and the variable w_3 then contains the magnitude of the filter DC offset. The background (Bgr),

calibration (Cal), offset (DC), linear slope (Lin), and all other vectors (spectra) with N elements can be interpreted as columns in the $N \times M$ matrix ϕ , where the M coefficients w_v are the weights, $v = 1, \dots, M$ (Fig. 10b):

$$\text{Amb}_{N \times 1} = \phi_{N \times M} W_{M \times 1} .$$

If we calculate the elements of the weight vector (w_1, \dots, w_M) from the matrix ϕ and the ambient signal (Amb), we obtain

$$W = (\phi^T \phi)^{-1} \phi^T \cdot \text{Amb} .$$

The value of w_2 represents, in this notation, the ratio of the ambient signal to the calibration signal and therefore can be used to determine the trace-gas concentration. The diagonal elements Cov_{ii} in the covariance matrix $\text{Cov} = (\phi^T \phi)^{-1}$ give the confidence range Δw_v for the M weights w_v .

This section describes a single-input filter that operates on sequential samples of the same signal. In contrast, the multiple linear regression scheme is a multiple-input filter with different signal sources, which uses the adaptive synthesis technique for the signal S_n . A representative of our scheme is the adaptive linear combiner in the form of a single-input adaptive transversal filter. This filter has only two input parameters, the ambient spectrum and the calibration spectrum. The example in Fig. 11 is a representation of the SSAT. An example for the PAT would involve exchanging the two spectra. The *adaptive Wiener filter*^{13,14} uses a MMSE algorithm to determine the optimum set of filter coefficients for a finite-impulse-response (FIR) filter. For each ambient spectrum a calculation of the filter coefficients is performed, and therefore the filter is able to adapt to changing conditions such as signal drift and changing background structures. For the Wiener filter a synthetic signal is calculated from the convolution of the calibration signal (Cal) and the $M = 2m + 1$ filter coefficients w_v :

$$S_n = \sum_{v=-m}^m w_v \cdot \text{Cal}_{n+v} ,$$

where $N \geq M$, i.e., the number of filter coefficients should be less than the number of channels in the input vector. Then the performance function is

$$\xi = \sum_{n=1}^N \left(\text{Amb}_n - \sum_{v=-m}^m w_v \cdot \text{Cal}_{n+v} \right)^2 .$$

Taking the partial derivatives of ξ with respect to the $2m + 1$ filter coefficients w_μ , solution for the minimum in the performance surface leads to

$$\frac{\partial \xi}{\partial w_\mu} = 2 \sum_{n=1}^N \left(\text{Amb}_n - \sum_{v=-m}^m w_v \cdot \text{Cal}_{n+v} \right) (-\text{Cal}_{n+\mu}) = 0 ,$$

or

$$\sum_{n=1}^N \text{Amb}_n \cdot \text{Cal}_{n+\mu} = \sum_{v=-m}^m w_v \sum_{n=1}^N \text{Cal}_{n+v} \cdot \text{Cal}_{n+\mu} .$$

If we define the input autocorrelation matrix

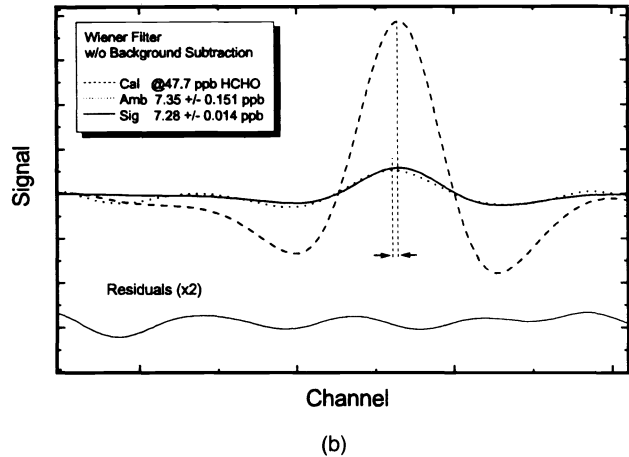
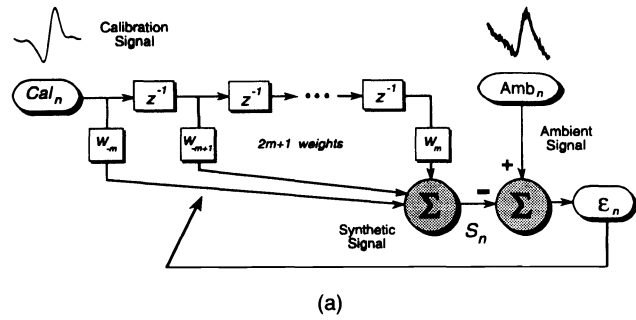


Fig. 11 (a) Adaptive linear combiner as single-input adaptive transversal filter. Time-varying coefficients represent an implementation of the Wiener filter. (b) Application of a Wiener filter to ambient signals.

$$\Phi_{v\mu} = \sum_{n=1}^N \text{Cal}_{n+v} \cdot \text{Cal}_{n+\mu}$$

and the cross-correlation vector

$$\gamma_\mu = \sum_{n=1}^N \text{Amb}_n \cdot \text{Cal}_{n+\mu} ,$$

we obtain $2m + 1$ simultaneous equations, which are an expression of the Wiener-Hopf equation

$$\gamma_\mu = \sum_{v=-m}^m w_v \Phi_{v\mu} .$$

The optimal weight vector W^* , which is sometimes called the Wiener weight vector, is

$$W^* = (\Phi^T \Phi)^{-1} \Phi^T \gamma .$$

The structure of the solution is the same as with the multiple linear regression scheme, but, due to the auto- and cross-correlation terms, this filter is able to handle signal drifts using sequential delays of z^{-1} [Fig. 11(a)]. Therefore, it can be interpreted as an efficient combination of the previously discussed on-line shifter with a subsequent regression when applied to low-pass filtered signals on a scan-by-scan basis. If, for practical reasons, the Wiener filter is applied to coaveraged spectra, only the overall drift on the integrated spectra

can be eliminated. The selection of the optimum procedure depends for example on the number of channels per spectrum and the number of filter coefficients, which determine the signal-processing time and the resulting duty cycle.

An application of a Wiener filter to ambient signals without background subtraction prior to signal processing is shown in Fig. 11(b). The spectrometer has been calibrated at 47.7 ppb formaldehyde from a stable permeation device. The synthetic sample showed a concentration of 7.35 ± 0.151 ppb. After the Wiener filter algorithm had been applied to the raw spectra, the data analysis reported a concentration of 7.28 ppb with an error of only 14 ppt. This indicates a significant improvement of about one order of magnitude in the confidence range of the data by adapting the ambient signal to the calibration signature. The resulting residuals, which have been derived from the filtered and the unfiltered spectra, show a periodic structure, as would be expected from a residual fringe structure. This filter has the interesting feature that besides ambient and calibration spectra no further information is, in principle, needed.

4 Signal Normalization

In most spectroscopic measurements the detected signal, which is proportional to the trace-gas concentration, depends on the laser power. Intensity fluctuations of the laser or changes in the optical-system alignment can generate a variation in the reported concentration data that is not real. Therefore, an intensity normalization has to be applied to measured data.

The principal spectroscopic setup for absorption measurements is shown in Fig. 12. The laser intensity I_0 , can be measured by coupling a certain amount of energy from the main beam, using a beam splitter, and monitoring the laser power drift with a reference detector. A more direct method is to monitor the average detector current i_D , which is (after dark-current subtraction) proportional to the number of photons incident on the detector. Figure 13 shows time-series data from 10-ppb formaldehyde as the calibration gas from a permeation source, and the associated errors from the fit. An Allan variance⁷ analysis of these raw data predicts a detection limit of 45 ppt and a maximum integration time of about 70 s. As the detector bias current was recorded simultaneously, the raw data have been normalized to the detector bias current (Fig. 14). The corresponding Allan plot indicates a lower maximum integration time than without normalization. This is due to the well-known fact that the deviation in the detector bias can be attributed to changes in the background radiation (i.e., sunset during these measurements) rather than changes in laser intensity or alignment.

A prerequisite for an improved normalization scheme, which will be described in this section, is the application of a modulation technique. The general idea is based on the measurement of the average modulation signal amplitude as shown in Fig. 15. Especially in the case of high-frequency modulation techniques, the rf power level at the detector ideally represents variations in laser power and system alignment changes due to slow drifts or vibrations.

When high-frequency modulation techniques are applied to a semiconductor diode laser, modulation frequencies around 100 MHz are usually used. At a typical FM modulation index, $\beta \approx 1$, a dominating upper and lower sideband with opposite phase are generated, which are separated by

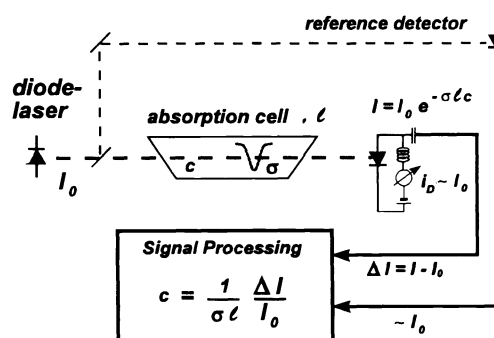


Fig. 12 Absorption spectroscopy with tunable diode lasers. Intensity normalization can be performed by using a reference channel or the background corrected detector bias current i_D .

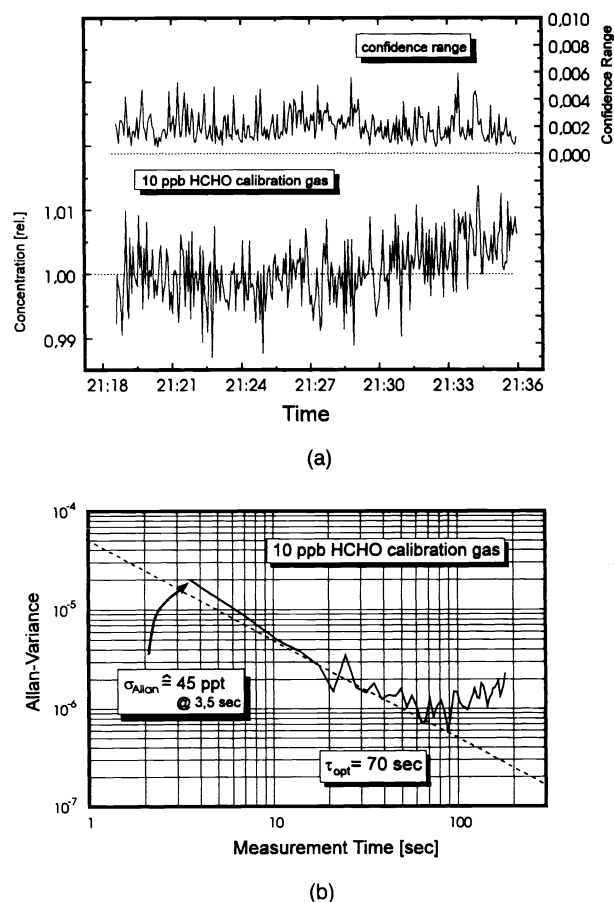


Fig. 13 Raw data from a 10-ppb formaldehyde permeation source with confidence range calculated by the fitting algorithm. The corresponding Allan variance analysis indicates a detection limit of about 45 ppt and an optimum integration time of 70 s.

$\pm \omega_{rf}$ from the laser carrier. At a square-law detector with a high bandwidth, harmonic mixing products of the modulation frequency will be generated, and due to the above-mentioned opposite phase of the upper and lower sidebands, the corresponding detector currents at the modulation frequency show a phase shift of 180 deg, so that the two currents cancel each other out. If the laser is tuned across a molecular absorption line, the sideband that overlaps the absorption line

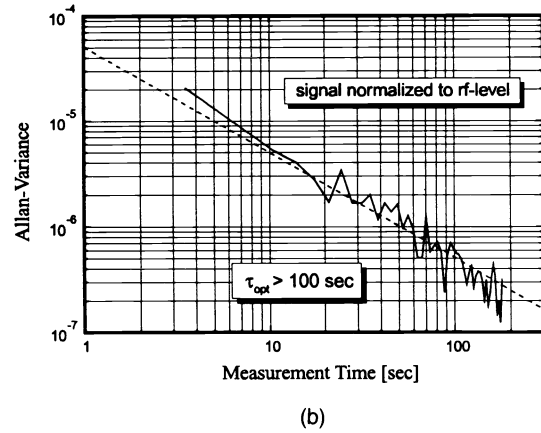
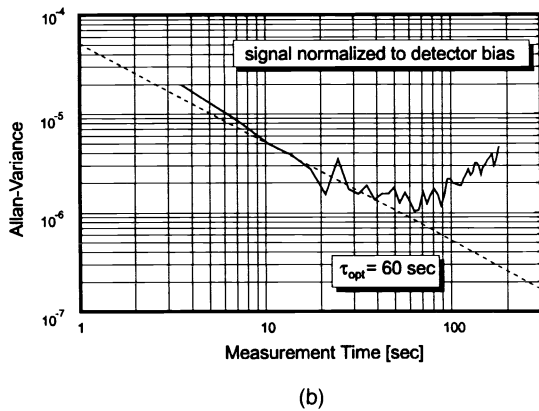
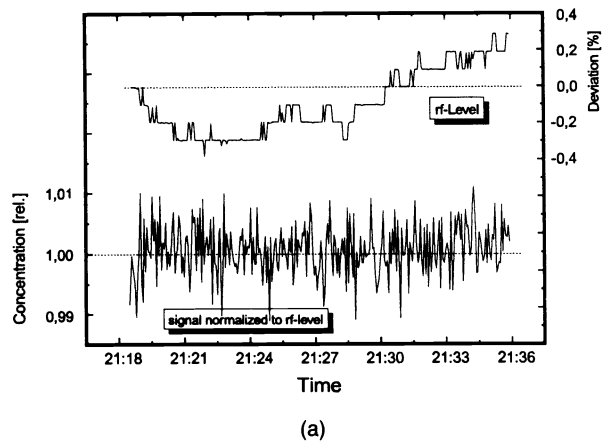
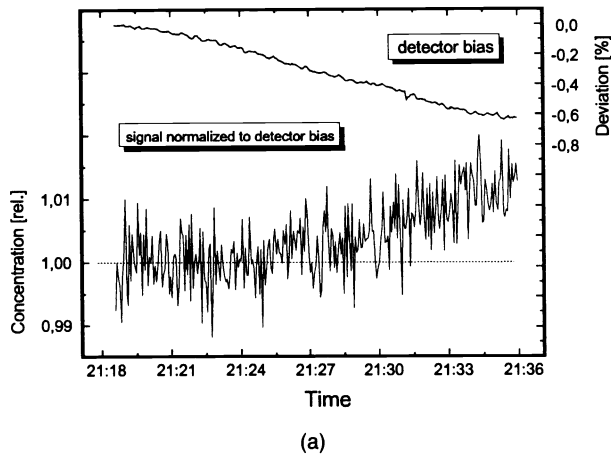


Fig. 14 Normalization of raw data from Fig. 13 to detector bias current and the corresponding Allan plot of normalized relative concentration data. Due to dominating changes in the background radiation, a significant degradation of system performance can be observed.

Fig. 16 Normalization of raw data to the change in the rf power level measured at the detector output, and the corresponding Allan plot of the corrected relative concentration data.

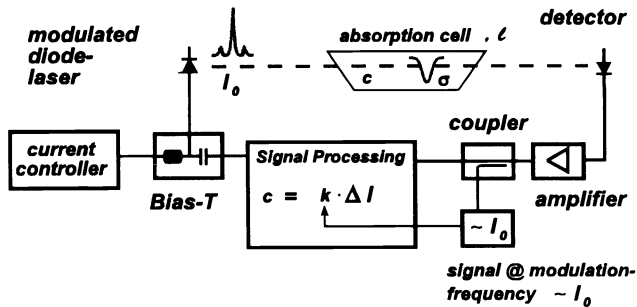


Fig. 15 Implementation of the rf intensity normalization in modulation spectroscopy.

will be attenuated, and therefore the *a priori* perfect balance is disturbed and a nonzero signal will be observed. Generally a coupling between amplitude and frequency modulation can be observed with semiconductor diode lasers.¹⁵ If carriers are injected periodically into the *pn* junction of the laser, the carrier density (amplitude modulation) and the index of refraction (frequency modulation) in the laser resonator change with the injection current. The degree of coupling of the two modulations depends on the laser operating point and struc-

ture, and manifests itself as the so-called residual amplitude modulation.¹⁶

With pure amplitude modulation the upper and lower sidebands are always in phase, and therefore the perfect cancellation of the zero signal, as mentioned above, cannot be attained. There will always be a residual rf current at the modulation frequency. This rf current is strictly proportional to the laser power incident upon the detector and accounts for changes in the laser intensity as well as for changes in the optical system alignment. As is shown in Fig. 15, a fraction of this current can be coupled out and used for intensity normalization. A normalization of the raw spectrometer data to the simultaneously recorded rf level at the detector (Fig. 16) leads to a significant improvement in system stability, and an optimum integration time of more than 100 s can be obtained, as can be seen from the corresponding Allan plot. In contrast to the conventional normalization techniques, the rf-normalization technique is insensitive to changes in the background radiation.

5 Summary and Conclusions

While many improvements in modern TDLAS systems focus on optimizing electronics and optical components, much less effort has been put into postdetection signal processing and

adaptive control. It is the purpose of this paper to discuss a signal-processing concept for applications of tunable-diode-laser spectroscopy and real-time signal analysis using sophisticated data-processing techniques.

It has been shown that spectroscopic signals vary in frequency due to fluctuations or transients in the laser current, and they vary in amplitude due to changes in background radiation and structures such as moving fringes. Broad fringes can appear as a linear slope superimposed on the desired TDLAS signal. The linear regression scheme conventionally used has been shown to report incorrect concentrations when the signal is distorted, which is especially critical at low ambient concentrations. Depending on the direction of the slope of the background, the linear regression tends to over- or underestimate the concentration of the ambient gas. Much more critical is the effect of drifts on the ambient spectra relative to the calibration spectrum. It is very important to note that these drift effects can lead to a significant *underestimation* of concentration values, i.e., a spectrometer that shows critical drifts can report extremely low concentration values, and so possible variations in calculated data may refer to drift effects rather than to changes in ambient trace-gas concentrations. Drift and slope effects have been investigated for symmetric signals (2f, two-tone FM) and for asymmetric signals (1f, single-tone FM). While the sensitivity towards drift is the same for both types of signals, it can be shown that perfectly symmetric signals are less susceptible to changes in the slope of the background.

To cope with the problems mentioned above, a signal shifter has been presented, which performs an on-line shift correction and thereby improves significantly the confidence range of the data. To reduce the sensitivity to changes in the background structure, multiple linear regression has been found to be useful, because linear, quadratic, and even higher-order residual structures due to changes in the background signal can be accounted for. Variations in the signal amplitude can be corrected by selection of a proper normalization scheme. A novel rf-normalization technique has been described, which accounts for changes in the laser intensity and changes in the optical alignment due to drifts and vibrations. In contrast to other normalization schemes, this technique is also completely independent of changes in the background radiation.

An adaptive control structure that has proved to be extremely useful is the adaptive Wiener filter. It has many applications in geophysics,¹⁷ multispectral imaging,¹⁸ and target detection.¹⁹ Usually targets must be detected against a spatially and temporally structured background. Often this background clutter²⁰ represents the noise source that limits the detectability of the target. To minimize the effect of clutter, complicated detection techniques must be utilized. The problem of clutter suppression in target recognition systems or infrared search-and-track (IRST) systems, where a target is within a time-varying background structure, is closely related to the problems in ultrasensitive spectroscopy, where a weak signal is covered by noise and interfering signals such as moving fringes. The purpose of an optimal filter is to recover the desired signal from the actually recorded signal. While the multiple linear regression scheme has no time delays, the Wiener filter is able to detect signal drifts and therefore has the "shift option" as an integral part. The most exciting feature of the Wiener filter is that, in principle, no

background spectrum is required and therefore background spectra could probably be omitted. Current work is focused on the implementation of on-line, adaptive Wiener and Kalman^{21,22} filters to improve system performance.

The application of modern filter and signal processing is a promising prospect and can help to improve precision and accuracy in tunable-diode-laser spectroscopy. This is of great importance in the context of quality control and quality assurance (QA/QC) for trace-gas measurements.

Acknowledgments

This work has been funded by the German Ministry for Research and Technology (BMFT) and the Bayerische Staatsministerium für Wirtschaft und Verkehr.

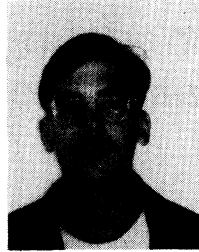
References

1. B. J. Finlayson-Pitts and J. N. Pitts, Jr., *Atmospheric Chemistry*, Wiley, New York (1986).
2. R. Grisar, G. Schmidtke, M. Tacke, and G. Restelli, Eds., *Monitoring of Gaseous Pollutants by Tunable Diode Lasers*, Kluwer Academic Publishers, Dordrecht, Holland (1992).
3. P. Warneck, *Chemistry of the Natural Atmosphere*, Academic Press, London (1988).
4. G. C. Bjorklund, "Frequency-modulation spectroscopy: A new method for measuring weak absorptions and dispersions," *Opt. Lett.* **5**, 15–17 (1980).
5. P. Werle, F. Slemr, M. Gehrtz, and Chr. Bräuchle, "Quantum-limited FM-spectroscopy with a lead-salt diode laser," *Appl. Phys. B* **49**, 99–108 (1989), and literature cited there.
6. D. S. Bomse, A. C. Stanton, and J. A. Silver, "Frequency modulation and wavelength modulation spectroscopies: Comparison of experimental methods using a lead-salt diode laser," *Appl. Opt.* **31**, 718–731 (1992).
7. P. Werle, R. Mücke, and F. Slemr, "The limits of signal averaging in tunable diode laser absorption spectroscopy," *Appl. Phys. B* **57**, 131–139 (1993).
8. H. Riris, C. B. Carlisle, R. E. Warren, and D. E. Cooper, "Signal-to-noise enhancement in frequency-modulation spectrometers by digital signal processing," *Opt. Lett.* **19**, 144–146 (1994).
9. P. Werle, "Signal processing strategies for tunable diode laser spectroscopy," in *Optical Sensing for Environmental Monitoring*, Proc. SPIE **2112**, 19–30 (1994).
10. K. J. Aström, "Theory and application of adaptive control—a survey," *Automatica* **19**, 471–486 (1983).
11. P. Werle, R. Mücke, and F. Slemr, "Development of a prototype IR-FM absorption spectrometer: Design criteria and system performance," in *Monitoring of Gaseous Pollutants by Tunable Diode Lasers*, R. Grisar, G. Schmidtke, M. Tacke, and G. Restelli, Eds., Kluwer Academic Publishers, Dordrecht, Holland (1992).
12. A. Fried, B. Henry, and J. R. Drummond, "Tunable diode laser measurements of atmospheric constituents by employing dual fitting analysis and jump scanning," *Appl. Opt.* **32**, 821–827 (1993).
13. N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, Wiley, New York (1949).
14. S. Treitel and E. A. Robinson, "The design of high resolution digital filters," *IEEE Trans. Geosci. Electron.* **GE-4**, 25–38 (1966).
15. M. Osinsky and J. Buus, "Linewidth broadening factor in semiconductor lasers—an overview," *IEEE J. Quantum Electron.* **OE-23**, 9–29 (1987).
16. M. Gehrtz, W. Lenth, A. T. Young, and H. S. Johnston, "High frequency modulation spectroscopy with lead salt diode laser," *Opt. Lett.* **11**, 132–134 (1986).
17. M. Bath, *Spectral Analysis in Geophysics*, Elsevier Science, New York (1974).
18. M. L. Althouse and C. I. Chang, "Chemical vapor detection with a multispectral thermal imager," *Opt. Eng.* **30**(11), 1725–1733 (1991).
19. D. Wang, "Adaptive spatial/temporal/spectral filters for background clutter suppression and target detection," *Opt. Eng.* **21**(6), 1033–1038 (1982).
20. W. L. Wolfe and G. J. Zisis, Eds., *The Infrared Handbook*, ERIM, Ann Arbor, MI (1989).
21. H. W. Sorenson, "Least-squares estimation: From Gauss to Kalman," *IEEE Spectrum* **7**, 63–68 (1970).
22. S. A. Billings and W. S. F. Voon, "Least squares parameter estimation for non-linear systems," *Int. J. Systems Sci.* **15**, 601–615 (1984).



Peter Werle works as a section manager at the Fraunhofer Institute for Atmospheric Environmental Research in Garmisch-Partenkirchen, where he is responsible for the development of integrated systems for trace-gas analysis. He received the Diplom-Physiker degree from the University of Mainz. His PhD thesis concerned high-resolution IR spectroscopy and was conducted at the Institute of Physical Chemistry of the University of Munich, where he

received his Dr. rer. nat. degree in 1989. His current research interests focus on improving the sensitivity and detection speed of spectroscopic instrumentation for trace-gas analysis in ambient air, ultrasensitive FM spectroscopy using semiconductor diode lasers, and digital signal-processing techniques. Dr. Werle is a member of the Deutsche Physikalische Gesellschaft, the Optical Society of America, and SPIE.



Bodo Scheumann received his Diplom Physiker degree from the University of Munich, Germany, and is currently working on his PhD thesis at the Fraunhofer Institute for Atmospheric Environmental Research in Garmisch-Partenkirchen, Germany. His work is focused on multicomponent ambient trace-gas measurements in the atmosphere using high-frequency modulation techniques applied to lead salt semiconductor diode lasers.



Josef Schandl received his Diplom Ingenieur degree in electrical engineering from the Fachhochschule München, Germany. He is working at the Fraunhofer Institute for Atmospheric Environmental Research in Garmisch-Partenkirchen, where he has developed a transputer-based platform for an automated tunable-diode-laser spectrometer and implemented the software for various control and signal-processing algorithms.