REVIEW ARTICLE

Exponential analysis in physical phenomena

Andrei A. Istratov[a] and Oleg F. Vyvenko
Institute of Physics of St.-Petersburg State University, Ulianovskaya 1, St. Petersburg 198904, Russia

(Received 13 July 1998; accepted for publication 16 November 1998)

Many physical phenomena are described by first-order differential equations whose solution is an exponential decay. Determining the time constants and amplitudes of exponential decays from the experimental data is a common task in semiconductor physics (deep level transient spectroscopy), biophysics (fluorescence decay analysis), nuclear physics and chemistry (radioactive decays, nuclear magnetic resonance), chemistry and electrochemistry (reaction kinetics) and medical imaging. This review article discusses the fundamental mathematical limitations of exponential analysis, outlines the critical aspects of acquisition of exponential transients for subsequent analysis, and gives a comprehensive overview of numerical algorithms used in exponential analysis. In the first part of the article the resolution of exponential analysis as a function of noise in input decays is discussed. It is shown that two exponential decays can be resolved in a transient only if the ratio of their time constants is greater than the resolution limit, which can be explicitly calculated from the signal-to-noise ratio in the transient. Although the signal-to-noise ratio is generally limited by the sensitivity of the equipment, it is shown that digitalization of the decays may be a major source of noise. The requirements for type of analog-to-digital converter, number of digitized data points and duration of digitized transients, which must be met to obtain the theoretical resolution limit and to improve stability of the exponential analysis, are formulated. The second part of the review article gives an overview and comparison of major numerical techniques of exponential analysis, such as the nonlinear least squares fit, the Prony method, the method of modulating functions, the method of moments, the Laplace–Padé approximation, the Tikhonov regularization method, the Gardner transformation, the method of maximum entropy and others. [S0034-6748(99)04502-5]

I. INTRODUCTION: EXPONENTIAL RELAXATION IN PHYSICS

First-order differential equations are among the most common equations in physics. They apply when the rate of increase or decrease of a certain function of time $f(t)$ is proportional to the value f itself:

$$\frac{df(t)}{dt} = -\lambda f(t). \tag{1}$$

The solution of this equation is an exponential decay of the form

$$f(t) = A \exp(-\lambda t) + B, \tag{2}$$

where A is the decay amplitude, B is a constant (the baseline offset), and λ is the decay rate (also called decay constant, or rate constant), which is inversely proportional to the time constant of the decay τ, $\lambda = \tau^{-1}$. Examples of application of Eqs. (1) and (2) are analysis of radioactive decays, ultrarelativistic heavy-ion experiments in nuclear physics, photon correlation spectroscopy, studies of fluorescent decays in biophysics, studies of sedimentation equilibrium, nuclear magnetic resonance in chemistry and medical imaging, and determination of molecular size distributions from laser light scattering data, just to mention a few. Exponential decays are common in solid state physics,[1] medicine,[2–6] biology and biophysics,[7–16] geophysics[17,18] optics,[19] engineering,[20] chemistry and electrochemistry.[21–28]

The amplitude of the exponential decay A and the decay rate λ carry information about the nature of the phenomenon being studied. The amplitude A usually corresponds to the initial concentration of the decaying species, and the decay rate λ is frequently determined by the energy change involved in the transition from one state of the system into another. Analysis of exponential decays given by Eq. (2) is straightforward and would not, by itself, deserve a special discussion. However, it commonly happens in physics that a number of exponential processes take place simultaneously, and experimental equipment yields a signal which is a sum of several exponential components. In this case, one has to solve a mathematical problem of decomposing a multiple exponential into its constituent parts. This task is not as

[a]Present address: University of California at Berkeley, LBNL, Mailstop 62-349, 1 Cyclotron Rd., Berkeley, CA 94720; electronic mail: istratov@socrates.berkeley.edu

simple as it may appear at first glance. In fact, it is one of the oldest and yet most persistent problems of functional analysis. As early as 1795, Prony[29] devised an algebraic scheme that could separate a small number of exponentials with similar amplitudes but substantially different time constants. The method, however, did not differentiate well between exponentials with close time constants. Prony recognized this limitation as fundamental, and many modern authors[30,31] have echoed this view.

The authors of this review article have been involved in investigations of electrical activity of defects in semiconductors[32,33] using the method of deep level transient spectroscopy (DLTS).[34] This method extracts information on deep level defects from the measurements of capacitance of the reverse-biased Schottky diode. The reverse bias is altered periodically, which results in capacitance transients. A major part of DLTS data evaluation is the exponential analysis of these transients. The increasing requirements for resolving closely spaced deep energy levels with similar decay time constants[35–37] stimulated the development of advanced numerical algorithms for DLTS data analysis. Our literature survey revealed that the problem of accurate exponential analysis is a very intensively studied problem that is common for many fields of science. A search in the INSPEC database using ''exponential analysis'' as a keyword returns as many as about 8900 articles published between 1969 and 1998. Unfortunately, in most cases researchers engaged in one area of science are not aware of similar studies made in the other fields. This results in parallelism in research and in an enormous wasted effort. Moreover, many physicists seem to be unaware of some fundamental limitations of the exponential analysis, which are known from the mathematical literature.

This review article is, to the best of our knowledge, the first article to summarize the mathematical and physical literature dealing with the problem of exponential analysis. Surely, we could not cover all areas of application of exponential analysis, and could not provide references to all articles where application of exponential analysis enabled experimentalists to make substantial progress in understanding physical phenomena. Instead, we present the readers with a general picture of the current status of exponential analysis, focused on fundamental understanding of problems involved in exponential analysis and on discussion of existing algorithms.

This article consists of three parts. In the first part (Secs. II–V) we discuss the principle limitations of exponential analysis and show that the resolution limit of the analysis is determined (and can be calculated using formulas reported in the literature) by the signal-to-noise ratio (SNR) in the exponential decays. Practical recommendations for data acquisition and averaging, which enable one to improve the accuracy of the exponential analysis, are given. In the second part (Sec. VI) we present an overview of major numerical algorithms for exponential analysis. Finally (Sec. VII), we compare these methods on the basis of literature data.

II. CLASSIFICATION OF PROBLEMS TO BE SOLVED IN EXPONENTIAL ANALYSIS

This review article deals with the numerical analysis of experimentally measured decaying functions of time $f(t)$ which stem from the processes described by the exponential law. However, $f(t)$ is not necessarily described by a single decay rate. In this treatment, we will emphasize three cases of exponential analysis. In the simplest case, further referred to as ''monoexponential analysis,'' the transient is assumed to be a single exponential, which is characterized by the decay amplitude A and decay rate λ:

$$f(t)=A\exp(-\lambda t). \tag{3}$$

The decay may also contain a baseline offset B as in Eq. (2). Most algorithms discussed in this review article require the baseline offset to be subtracted before analysis. Therefore, we will assume in the following (except where specified explicitly) that the baseline offset $B=0$.

If the decay consists of a sum of n exponentials of the form Eq. (3), we will discuss the analysis of ''multiexponentials decays,'' or ''multiexponential analysis:''

$$f(t)=\sum_{i=1}^{n} A_i\exp(-\lambda_i t). \tag{4}$$

The goal of the multiexponential analysis is to determine the number of exponential components n, their amplitudes A_i, and decay rates λ_i. Finally, in the general case when the decay is described by a continuous distribution of emission rates given by a spectral function $g(\lambda)$ rather than by a sum of discrete exponential transients, we will discuss the analysis of ''nonexponential transients,'' or spectroscopic methods of exponential analysis:

$$f(t)=\int_0^{\infty} g(\lambda)\exp(-\lambda t)d\lambda. \tag{5}$$

Analysis of nonexponential transients is aimed at determining the spectral function $g(\lambda)$. Equation (5) reduces to Eq. (4) if the spectral function $g(\lambda)$ can be presented as a sum of n delta functions:

$$g(\lambda)=\sum_{i=1}^{n} A_i\delta[\lambda-\lambda_i]. \tag{6}$$

In addition to the definitions of three types of exponential decays [Eqs. (3), (4), and (5)], the experimentally measured decays $f_{\exp}(t)$ contain a noise component $\epsilon(t)$: $f_{\exp}(t)=f(t)+\epsilon(t)$.

An example of all three cases, taken from solid state physics, is given in Fig. 1. The top of the figure represents an energy band diagram of a semiconductor with one [Fig. 1(a)], three [Fig. 1(b)] and multiple [Fig. 1(c)] trap levels. These traps can capture electrons and then emit them back to the conduction band. The emission process is described by exponentials with the time constants, depending on the temperature and the energy position of the traps in the band gap. The case of a single trap level corresponds to a monoexponential decay. The spectral function $g(\lambda)$, which determines the amplitude of exponential decay with the emission rate

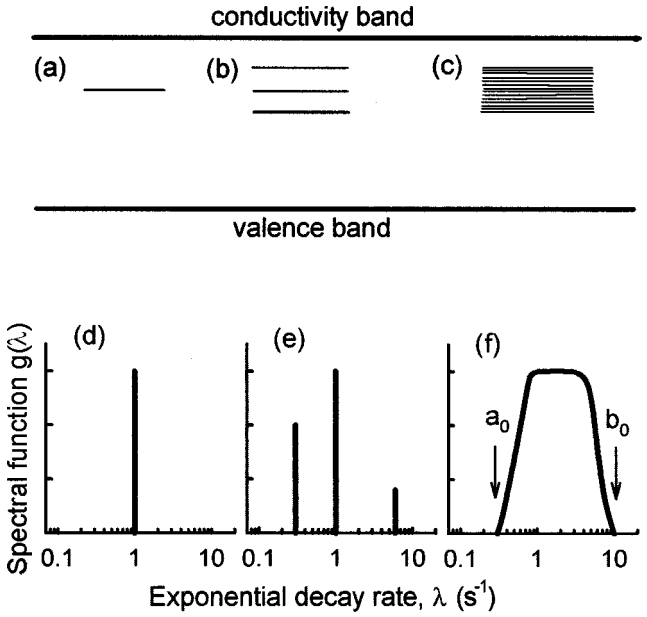

FIG. 1. A band diagram of a semiconductor with a single deep level (a), three deep levels (b) and a continuous energy distribution of levels of noninteracting defects (c), and corresponding to these three cases emission rate spectra of the decays: monoexponential decay (d), multiexponential decay (e) and nonexponential decay (f).

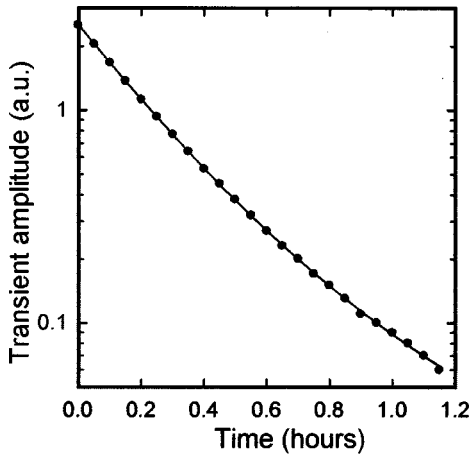

FIG. 2. The famous example of Lanczos (Ref. 30). Twenty four data points (filled circles) are fitted by a double exponential $f_2(t)=2.202\exp(-4.45t)+0.305\exp(-1.58t)$ (dashed line) and by a triple exponential $f_3(t)=0.0951\exp(-t)+0.8607\exp(-3t)$ and $1.5576\exp(-5t)$ (solid line). The difference between $f_2(t)$ and $f_3(t)$ is less than the line width, and the lines are undistinguishable. Following Ref. 30, the units of time are hours.

λ_0, is presented in Fig. 1(d) and is equal to zero everywhere except for $\lambda=\lambda_0$. The case of three trap levels results in decays, which consist of a sum of three exponentials. In this case the spectral function $g(\lambda)$ is represented by three delta function-like spikes, as in Fig. 1(e). Finally, the case of a distribution of noninteracting trap levels is described by a smooth continuous spectral function $g(\lambda)$ in Fig. 1(f). It is assumed that the function $g(\lambda)$ takes nonzero values over the interval $[a_0,b_0]$ and is zero for the emission rates outside of this interval. In the following we will call this interval a domain of the function $g(\lambda)$.

III. FUNDAMENTAL LIMITATIONS OF THE EXPONENTIAL ANALYSIS

The problem of exponential analysis is solved, in principle, by taking the inverse Laplace transform of the transient $f(t)$ (Ref. 38):

$$g(\lambda)=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} f(t)\exp(\lambda t)dt, \tag{7}$$

where c is a real constant. It is usually quite straightforward to compute the integral Eq. (7) if the analytical expression for $f(t)$ is known. This, however, seldom happens in experimental physics. In most cases, Eq. (7) (also known as a Bromwich integral) cannot be applied directly to experimental data. Unlike sinusoids, exponentials are not orthogonal along the real axis, i.e., the contribution of each exponential to the signal cannot be projected out by taking an inner product defined as an integral along the real axis. This is reflected by the fact that Eq. (7) requires integration in the complex plane. Yet, from experimental observations, only values of the signal function along the real axis are known. Thus, the spectral function $g(\lambda)$ can be found only by solving the Laplace integral equation, Eq. (5). This equation belongs to a more general class of Fredholm integral equations of the first kind, which are known to be ill posed[39,40] (or ''incorrectly'' or ''improperly'' posed). This term means that the solution, $g(\lambda)$, of Eq. (5) may not be unique, may not exist, and may not depend continuously on the data.

The ill-posed nature of the Laplace integral equation can be easily understood[41] if we take a Fourier transform of both sides of Eq. (5). The equation becomes:

$$\hat{f}(\omega)=\hat{K}(\omega)\times\hat{g}(\omega). \tag{8}$$

Here, $\hat{f}(\omega)$, $\hat{K}(\omega)$, and $\hat{g}(\omega)$ are the Fourier transforms of $f(t)$, $K(t,\lambda)$ and $g(\lambda)$, respectively. It can be shown[42,43] that the Fourier ''image'' of the Laplace kernel $K(t,\lambda)=\exp(-\lambda t)$ is band limited, i.e., $\hat{K}(\omega)$ decreases to zero as $1/(\omega^2+\lambda^2)$ for $\omega\to\infty$. The Laplace operator can thus be compared with a low-pass filter in electronics. Using this analogy, one can say that the high-frequency components of the Fourier spectrum: $\hat{g}(\omega)$ of $g(\lambda)$, are cut off by the band-limited Laplace integral operator if $\omega>\omega_{\max}$, where $\omega_{\max}$ is a certain threshold frequency. For instance, if $\hat{g}^{(0)}(\omega)$ is such a function that $\hat{g}^{(0)}(\omega)=0$ for $|\omega|<\omega_{\max}$, but $\hat{g}^{(0)}(\omega)\neq 0$ for $|\omega|>\omega_{\max}$ (for example, $g^{(0)}$ may represent high-frequency noise), then $\hat{K}(\omega)\times\hat{g}^{(0)}(\omega)=0$. If a solution $\hat{g}$ of Eq. (8) exists, then $\hat{g}+\hat{g}^{(0)}$ is also a solution, and thus the solution of the Laplace equation is not unique. This result is essentially the subject of the Riemann–Lebesgue theorem, well known from mathematics.[44,45] Since the Laplace kernel K acts as a smoothing operator which filters out high-frequency components of the function $g(\lambda)$, then any attempt to recover these components from a noisy transient $f_{\exp}(t)$ will lead to arbitrary deviations of $g(\lambda)$. Many physicists have discovered after much wasted effort that it is essential to understand the ill-conditioned nature of the problem before attempting to compute solutions. Examples which show how significantly two solutions g_1 and g_2 may differ

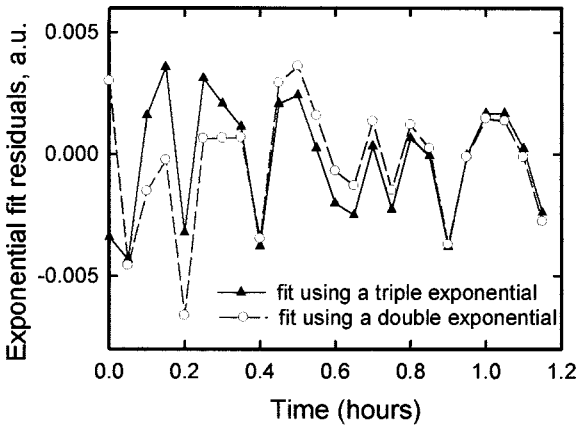

FIG. 3. Residuals (differences between the data points and the fit) calculated using the data and the fitting functions from the example of Lanczos (Ref. 30, Fig. 2).

were provided by Lanczos,[30] Julius[46] and Grinvald.[47] Lanczos[30] showed that a sum of two exponentials could be reproduced to within two decimal places by a sum of three exponentials with entirely different time constants and amplitudes. The example of Lanczos is reproduced in Fig. 2, where the experimental data (circles) and double- and triple-exponential fits (lines) are plotted. Although the plot contains two lines, they are undistinguishable since the difference between them is less than the linewidth. In numerical fits, residuals (the difference between the experimental data and the fit) are often used to evaluate the quality of the fit. Systematic deviation of the residuals from zero is considered as an indication of a poor choice of the fitting function. The residuals for the double- and the triple-exponential fits from Fig. 2 are presented in Fig. 3. Analysis of Fig. 3 shows that the residual plots for the double and triple exponentials look very similar and are not very helpful in deciding which fit is better.

It is natural to ask why the residual plot does not show any substantial difference between the two fits, although the number of exponential components and their decay rates differ quite significantly. The answer is that it is the consequence of nonorthogonality of exponentials. The variation of the weighted sum of squares of the residuals due to a change in the value of one parameter can be compensated to a considerable extent by adjusting the other parameters. The larger the number of parameters to be determined, the more serious this problem becomes. It is very important to realize that this last property is determined by the Laplace kernel K and is true also for noise-free decays $f(t)$.

Since the function $f_{\exp}(t)$ is measured experimentally, i.e., it contains noise and thus is not known accurately, then by solving the Laplace integral equation we get a family $\wp$ of functions $g(\lambda)$, which satisfy with prescribed degree of error the equation $L[\wp] \cong f_{\exp}(t)$, where L is the Laplace operator. The problem is then to pick the true solution $g_0(\lambda)$ out of the family $\wp$.

Fortunately, in many cases we have some prior information about the function $g_0(\lambda)$. This helps us to extract from the set of possible solutions a solution which has a physical meaning. The procedure of selecting a single solution from a set of possible ones is known as *regularization* of the integral equation. To regularize the equation, the inversion method must accommodate existing prior information. For example, a constraint $g(\lambda) > 0$ is very powerful for eliminating oscillating solutions. Another very important principle is the principle of parsimony which states that, of all possible solutions that have not been eliminated by prior knowledge, you must choose the simplest one, i.e., the one that reveals the least amount of detail or information that has not been already known or expected. While this solution may not have all the details of the true solution, the details which it has are necessary to fit the data and therefore less likely to be an artifact.[48] Usually, it is assumed that the ''simplest'' solution $g_0(\lambda)$ is the smoothest one. The smoothness of a function can be measured, for example, as an integral of its second derivative (Ref. 49) or as the amplitude of its high-frequency Fourier components. Another approach which can be used to find the solution to an ill-posed problem is the assumption that the solution has a predetermined form, for example, that it is a sum of n discrete exponential components [Eq. (4)]. Such a solution is known as a *quasisolution* in the sense of Ivanov.[50]

IV. THE RESOLUTION LIMIT OF EXPONENTIAL ANALYSIS

The problem of solving a Fredholm equation of the first kind [Eq. (5)] is not unique to exponential analysis. In much of experimental science, the data delivered by an experimental system are related to the phenomena under investigation by a linear integral transformation. The analysis of such systems has given rise to a well developed theory of ''resolution'' or ''information,'' associated with the names of Nyquist[51,52] and Shannon.[53,54] One finds that, in the presence of noise, the details of the ''object'' $g(\lambda)$ can only be recovered from the ''image'' $f(t)$ up to a certain resolution limit. This limit is characterized by the Shannon or Nyquist number, or Rayleigh criterion, and is determined by properties of the eigenvalue spectrum of the transformation $g \rightarrow f$. The eigenvalues γ_n and eigenfunctions ϕ_n are widely used in the theory of integral equations (see, e.g., Ref. 55). In information theory eigenfunctions are considered as basic elements of information which retain their identity under the action of the integral operator, but are scaled in magnitude by the eigenvalues γ_n:

$$\int_a^b K(x,y)\,\phi_n(y)\,dy = \gamma_n \phi_n(x). \tag{9}$$

A very important property of the eigenfunctions ϕ_n is that they form an orthogonal basis, and both the data function $f(t)$ and the solution $g(\lambda)$ can be represented as eigenfunction expansions:

$$f(t) = \sum_{k=1}^{\infty} f_k \phi_k(t), \quad g(\lambda) = \sum_{k=1}^{\infty} g_k \phi_k(\lambda). \tag{10}$$

Substituting Eq. (10) into the integral equation Eq. (5), then using Eq. (9) and the orthogonality of eigenfunctions ϕ_k, one can show that

$$g(\lambda)=\sum_{k=1}^{\infty}\frac{f_k}{\gamma_k}\phi_k(\lambda), \tag{11}$$

where it is assumed that the γ_k are arranged in descending order $\gamma_1>\gamma_2>\gamma_3\ldots$. The larger γ_k is, the more efficient the transmission of the corresponding information element ϕ_k is through the integral. Elements corresponding to small γ_k are transmitted so weakly that they become lost in noise and cannot be determined accurately. Unfortunately, as will be discussed below, the sequence of eigenvalues γ_k of the Laplace transform decreases to zero very quickly. Since the high-frequency noise components in the data, $f_{\exp}(t)$, prevent f_k in Eq. (11) from decreasing as fast as γ_k, the ratio f_k/γ_k diverges rapidly for large k. Thus, the calculations will eventually become noise limited and the series, Eq. (11), must be terminated. The function $g(\lambda)$ restored in such a way will contain the most information that can safely be recovered from the experimental data $f_{\exp}(t)$. The faster the sequence of γ_k decays to any given noise level, the sooner the series, Eq. (11), should be terminated, and the less information can be extracted from the raw data. It is important to realize that since the sequence of eigenvalues of the Laplace transform decays to zero, one can never obtain the exact solution to Eq. (5).[56] An infinite amount of information about the solution $g(\lambda)$ is in principal not recoverable from the measured transients $f_{\exp}(t)$.

The Fredholm equation with a Laplace kernel did not receive detailed consideration from the point of view of information theory until the article by Petrov *et al.*[57] In following articles, McWhirter *et al.*[56] and Pike *et al.*,[58] Bertero *et al.*[59–65] and Ostrowsky *et al.*[66] calculated the eigenfunctions and eigenvalues of the Laplace transform and identified the resolution elements of exponential analysis. Bertero *et al.*[60] defined the number of degrees of freedom, or generalized Shannon number M as the number of singular values that are greater than the SNR. The larger this generalized Shannon number is, the greater is the information capacity of the integral transform. McWhirter *et al.*[59] showed that the sequence of eigenvalues of the Laplace transform decays so quickly that in most cases only 4–9 eigenvalues of Eq. (5) are greater than a realistic SNR. For comparison, the eigenvalue spectrum of the Fourier transform never decays to zero,[56] which means that the Fourier transform has a much greater information capacity than the Laplace transform.

An important conclusion made by Ostrowsky *et al.*[66] is that the so-called sampling theorem known from Fourier analysis[54,67] can be applied to the problem of inversion of the Laplace transform. The sampling theorem states that the wave form is completely determined by its values at time intervals $1/2\omega_{\max}$, where $\omega_{\max}$ is the highest frequency present in its spectrum. Consequently, if the spectrum of the function $f(t)$ is band limited by the frequency $\omega_{\max}$, then all available information in the spectrum of $f(t)$ will be sufficient to restore $f(t)$ only in $1/2\omega_{\max}$ data points equidistantly spaced in t. Ostrowsky *et al.*[66] have shown that the sampling theorem applied to exponential analysis requires that the ''source'' $g(\lambda)$ be determined at equidistant points in logarithm of the emission rate, since the zeros of eigenfunctions of the Laplace transform are distributed approximately equidistantly on a logarithmic scale. The closest distance between exponential decay rates that can be resolved in exponential analysis in the case of an infinite domain of $g(\lambda)$ is given by:[59]

$$\delta=\lambda_i/\lambda_{i+1}=\exp(\pi/\omega_{\max}), \tag{12}$$

where $\omega_{\max}$ is determined by the SNR in the transients:

$$\cosh(\pi\omega_{\max})=\pi(\mathrm{SNR})^2. \tag{13}$$

Equations (12) and (13) determine the resolution limit of the exponential analysis. Attempts to increase the resolution by trying to determine $g(\lambda)$ at points closer than $\lambda_i/\lambda_{i+1}=\exp(\pi/\omega_{\max})$ are bound to yield unreliable results.[56,66]

It is important to note that prior knowledge about the domain of the solution partly compensates for the information lost in noisy decays, and can be used to achieve a further increase in resolution.[66] In Ref. 59 it was shown that the singular value spectrum of the Laplace integral operator decreases slower to zero if $g(\lambda)$ is defined on a finite interval $[a_0,b_0]$ [see Fig. 1(f)] as compared to the eigenvalue spectrum of a problem where $g(\lambda)$ is assumed to be defined on an infinite domain. With application to optics, this made it possible to obtain a resolution beyond the classical diffraction limit.[68] In the case of a limited domain of the function $g(\lambda)$, the resolution limit is determined by the number M of singular values which exceed the SNR:

$$\delta=\tau_i/\tau_{i+1}=\left(\frac{b_0}{a_0}\right)^{1/M}. \tag{14}$$

The resolution limit for infinite and finite domains is given in Table I and presented in Fig. 4. The values in Table I and Fig. 4 follow from Eqs. (12)–(14) and were obtained by Bertero *et al.*,[59] who calculated eigenfunctions and eigenvalues of the Laplace integral operator for different domains of the function $g(\lambda)$. Figure 4 shows that the resolution of exponential analysis can be substantially improved if the domain of the solution is known. If the domain of $g(\lambda)$ is not known *a priori*, one can use a ''zooming'' technique, i.e., begin with the assumption of an infinite domain, determine where the solution is localized, and then make another calculation for the estimated finite domain. Another method would be to estimate the localization of the solution from the knowledge of its first and second moments, which can be derived from the data before inversion.[69]

Summarizing this section, we came to the very important conclusion that there is a limit to the maximum resolution capacity of exponential analysis. This limit is inherent in the problem itself and is exacerbated by noise. This limit cannot

TABLE I. The resolution limit of exponential analysis for different signal-to-noise ratios in the input transients and for different domain of the solution $g(\lambda)$ (after Bertero *et al.*, Ref. 59).

Signal-to-noise ratio in transient $f(t)$	Domain of $g(\lambda)$, $\lambda_{\max}/\lambda_{\min}$		
	Infinite	5	2
10^2	2.44	1.74	1.44
10^3	1.88	1.45	1.27
10^4	1.63	1.32	1.20

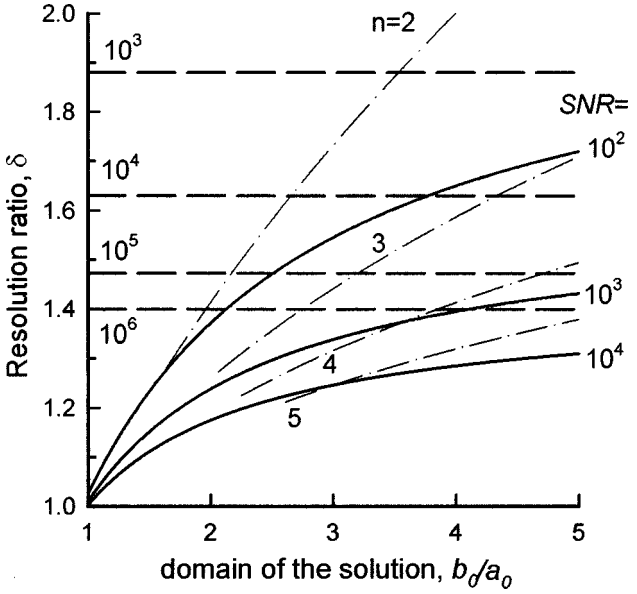

FIG. 4. The resolution limit of exponential analysis δ as a function of the domain of the solution b_0/a_0 and of the signal-to-noise ratio (SNR) (after Bertero *et al.*, Ref. 59). Solid lines represent dependencies of the resolution limit on domain width for different signal-to-noise ratios. The dash-dot lines give the resolution necessary to recover $n=2,3,4,5$ exponentials as a function of the domain; n exponential components can be resolved for a given SNR and domain b_0/a_0, when the dash-dotted line corresponding to n is below the solid line associated with SNR for the abscissa value of b_0/a_0. The horizontal dashed lines represent resolution limit for different SNR for the infinite domain.

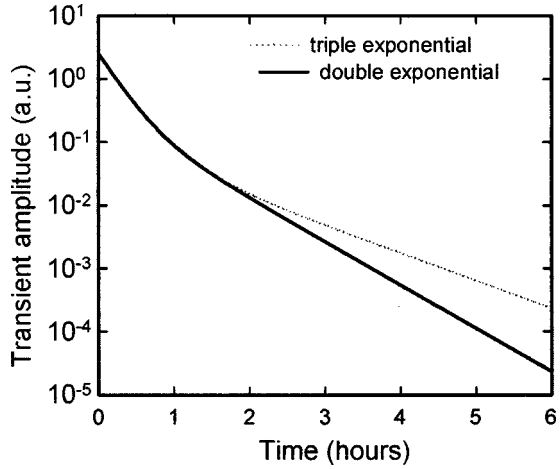

FIG. 5. The example of Lanczos (Ref. 30) plotted for longer decay times than in Fig. 2. It is obvious that the two curves, which are indeed undistinguishable at $t<2$ h, are well separated for $t>3$ h. However, the absolute value of the separation is less than only 0.001 of the decay amplitude.

be surmounted by numerical algorithms. Any attempt to obtain resolution beyond the resolution limit will result in unreliable and unstable solutions.

V. DATA ACQUISITION FOR HIGH-RESOLUTION EXPONENTIAL ANALYSIS

A. Digitalization and averaging of transients

The major goal of exponential analysis is to distinguish exponential components with close time constants in the experimentally measured decay. To achieve high resolution in exponential analysis, it is very important to record the transient until it decays completely.[70–72] Since the ratio of amplitudes of two exponentials with close decay rates: $\exp(-\lambda_1 t)$ and $\exp(-\lambda_2 t)$ increases with the time as $\exp[(\lambda_2-\lambda_1)t]$, then these exponentials always can, at least theoretically, be distinguished if the decay is monitored for a sufficiently long time. Since the exponential is a decaying function of time, the transient should be monitored as long as the signal amplitude exceeds the noise level. For a signal-to-noise ratio, SNR$=100$, the measurement time T should be at least 4.6τ [since $\exp(4.6)\approx 100$], for SNR$=1000$ about 6.9τ, and for SNR$=10^4$ at least 9.2τ. A too short duration T may become an important limiting factor for the resolution capacity. This is frequently ignored in experiments and numerical simulations. A number of examples confirming that the resolution of exponential analysis can be improved by increasing T/τ can be found in the literature. For instance, Smith *et al.*[73] analyzed the Gardner transform technique (described in Sec. VI C 4) and came to the conclusion that the best resolution in the recovered emission rate spectrum could be obtained only if the decay is known for at least 22τ. Hall *et al.*[74] compared the method of moments and the nonlinear least squares minimization and reported that the error in determination of τ decreased with the increase of measurement time and reached its minimum at T/τ in the range 10–16. Zhang *et al.*[75] reported that for a monoexponential transient with unknown baseline both the Prony (Sec. VI B 3) and Levenberg–Marquardt (Sec. VI B 2) algorithms require $T>5\tau$ for best performance.

To provide an example of the importance of monitoring the decays for the time periods which substantially exceed the decay time constants, we calculated the double- and triple-exponential decays from the example of Lanczos[30] (Sec. III) for the decay time of 6 h instead of 1.2 h as in Figs. 1 and 2. This corresponds to the increase of the T/τ ratio for the slowest component of the decay from 1.2 to 6. It is instructive to see in Fig. 5 how two decays, undistinguishable for $t<2$ h, become well separated after 3 h. However, the difference between the two curves does not exceed 0.001 of the decay amplitude, and hence can be detected only if the SNR in the experimental data exceeds 1000.

As discussed above, SNR in the input decays is a major factor that limits the resolution of exponential analysis. Unfortunately, SNR of the decays obtained from the experiment is usually determined by sensitivity of the equipment and often does not exceed 100. For example, interfering radio frequencies, instabilities of pulse generators and dc voltage sources, noise of current and voltage amplifiers, drift of temperature, or in optical spectroscopy fluctuations of lamp intensities, photomultiplier color effects, or light scattering, will all degrade the SNR. The above mentioned value of SNR$=100$ is still too low for most of algorithms discussed in Sec. VI, and is clearly insufficient to distinguish the curves in Fig. 5. Averaging a large number of transients enables one to improve the SNR by a factor of about $K^{1/2}$, where K is the number of averaged transients (this estimate assumes Poisson statistics). Since it is usually too time consuming to average more than about 10^4 transients, it is easy to calculate that after averaging one can obtain a SNR up to

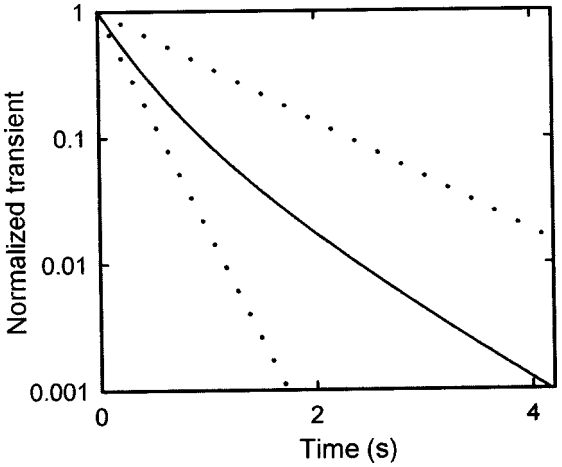

FIG. 6. Simulated transient obtained by averaging exponential decays with decay rate gradually changing from one to four (solid line). The dotted lines represent the fastest and the slowest components of the decay.

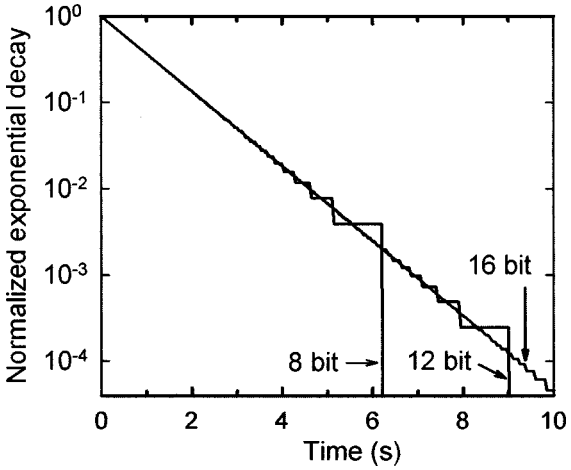

FIG. 7. A simulated exponential decay after digitalization using 8, 12, and 16 bit of the ADC. Steps on the curves are associated with digitalization noise.

10^4 for SNR without averaging of 100. According to Table I, this SNR would enable one to obtain the resolution as high as $\tau_1/\tau_2 \approx 1.63$. Such a resolution was indeed obtained in practice (see Ref. 72, for an example). However, one should keep in mind that it is necessary to avoid any fluctuations and drifts of the components of experimental setup during accumulation of transients. The averaged transients can be substantially distorted by instabilities of gain or a dc offset of amplifiers, by drifts of voltage and current sources or pulse generators, instabilities of temperature or lamp intensity, especially for long measurements. Although the amplitude of high-frequency noise components will decrease with increasing accumulation, the accuracy of the measured decay may actually start to decrease after a certain critical accumulation time is exceeded. For example, according to Dobaczewski,[76] the maximum averaging time of capacitance transients in high-resolution DLTS, which is tolerable to the drifts in the equipment, is about 10 min. This value may indeed depend upon the experimental equipment and the skill of the experimenter and may vary for different areas of science. It should be emphasized that the drifts which may change the decay time constant are particularly detrimental for exponential analysis since the average of several exponentials with different time constants is not a monoexponential decay. To illustrate this point, in Fig. 6 we present results of a simulation of averaging of a sequence of exponential decays with the decay rate changing gradually from unity to four. Dotted lines correspond to the fastest and the slowest decays of the set. The solid line is the averaged curve. Obviously, the accumulated decay is no longer monoexponential and does not properly represent the physical phenomenon which causes the decay.

An important part of an automated data acquisition system is an analog-to-digital converter (ADC). Important parameters of the ADC are its linearity and time stability. Furthermore, an ADC may introduce additional discretization noise[77] with an amplitude that can be roughly estimated as the sensitivity of the lowest bit of the ADC (i.e., as a change in the input voltage which changes readings of the ADC from N to $N+1$). It is very important for high-resolution exponential analysis that the amplitude of the input signal covers the full dynamic range of the ADC. This can be usually achieved by adjusting the sensitivity of the ADC (most modern data acquisition boards have input amplifiers with adjustable gain; if an on-board amplifier is not available, an external voltage amplifier can be used). If the amplitude of the transient is, e.g., only a hundred times larger than the lowest bit of the ADC, then the signal-to-noise ratio in the digitized transient will never exceed a level of about 100. Only 16-bit or, in the worst case, 12-bit ADCs are suitable for high-resolution exponential analysis. Figure 7 is an example of digitalization noise. In this figure, we simulated digitalization of an exponential transient with the amplitude of 20 meV using a 16-bit ADC with the sensitivity selected in such a way that the input voltage range is from 0 to 20 meV, 320 meV, and 5.12 V. The input signal is then spanned over the range corresponding to 16, 12, and 8 bits of the ADC. Steps on the curves in Fig. 7 correspond to the digitalization noise, which is fairly low if all 16 bits of the ADC are used, but becomes almost as high as 1% of the decay amplitude if the dc input voltage range is set much higher than the signal amplitude.

Another question that arises in experiments is how many experimental points in a transient should be measured for a high-resolution analysis. As was discussed in Sec. IV, the Laplace transform is band limited and it is only possible to derive a finite amount of information on the function $g(\lambda)$ from the data. The minimum number of data points necessary to determine $g(\lambda)$ for a given SNR is determined by the number of degrees of freedom of the solution, which in the case of inverse Laplace transform seldom exceeds 10 (see Sec. IV). This means that in most cases only 10–20 points would be sufficient. However, these points should be optimally (equidistantly in logarithmical scale) sampled.[66] Bertero *et al.*[60] showed that in the case of a limited domain $[a_0, b_0]$ of $g(\lambda)$, $b_0/a_0 \leqslant 8$, the ill conditioning of the restoration of 2–4 exponential components with 5 optimally placed data points is less than with 32 linearly spaced data samples.

Experience with well-posed problems tells us that the

more data points are measured on the curve, the more accurately it can be fitted. In exponential analysis one should take into account that an excess of data points makes the inversion problem less stable. The procedure of digitalization of the exponential decay is a kind of low-pass filter. In agreement with the Shannon sampling theorem,[54] only the oscillatory components with a frequency lower than half of the digitalization frequency are transferred. If we increase the sampling rate, we increase the content of noise-related high-frequency Fourier components in the spectrum of the measured transient. Since useful information can be extracted only from the several lowest-frequency Fourier components of the exponential decay that are greater than the noise, the additional noise-related high frequency components make the problem more ill posed.[60,78] Most regularization techniques discussed in the second part of this review article enable the user to find a stable solution even for a very large number of data points. However, one should realize that ultimately the increasing number of data points results in an increase of computational time without any improvement in resolution.

There is, however, another aspect of the same problem. As a rule, ADCs sample the transient at equidistant points of time. Since the sampling interval should be at least several times shorter than the time constant of the fastest component of the exponential decay, then for a spectroscopic analysis covering several orders of magnitude in τ, one has to measure from several thousands to several hundreds of thousands of points. In this case, one has to deal with very large data files. To reduce the number of data points in the analysis several authors suggested pseudologarithmic storage[79] or data acquisition[80] algorithms.

B. Extraction of the baseline offset

Determination and extraction of the baseline offset, B in Eq. (2), is an important problem of exponential analysis.[81] As it will be discussed in Sec. VI, most algorithms cannot accommodate the baseline offset as a parameter and require that it be extracted before the analysis. The existing methods of baseline offset subtraction are based on an assumption that even in the case of a multiexponential transient the tail of the decay can be approximated by a single exponential of the form:

$$f_0(t)=A\exp(-\lambda_0 t)+B. \tag{15}$$

Kirchner *et al.*[82] and Ikossi-Anastasiou *et al.*[83] suggested using the Fourier transform to determine the parameters A, B and λ_0. The formulas used in their method are described in detail in Sec. VI A 1.

Moore *et al.*[84] and Smith *et al.*[85] suggested two algebraic methods. In the method of Moore *et al.*[84] the value of the baseline offset B is treated as a parameter to be adjusted to minimize the variance of the distribution of ν_k determined as

$$\nu_k=\{\ln[f(t_k)-B]-\ln[f(t_{k+N/2})-B]\}/(t_k-t_{k+N/2}). \tag{16}$$

The value of k varies in Eq. (16) from 1 to $N/2$, where N is the number of experimental points.

Smith *et al.*[85] proposed to determine the baseline using three data points. If:

$$\begin{aligned} Y_1&=A_0\exp(-t_0/\tau)+B,\\ Y_2&=A_0\exp[-(t_0+t_1)/\tau]+B,\\ Y_3&=A_0\exp[-(t_0+2t_1)/\tau]+B \end{aligned} \tag{17}$$

then

$$B=(Y_1Y_3-Y_2^2)/(Y_1+Y_3-2Y_2). \tag{18}$$

Alternatively, the decay can be differentiated to remove the baseline.[70,86] This can be done by taking the differences between the values of the signal at successive points and carrying out an exponential analysis of these differences instead of the original decay.[87] Mangelsdorf[88] and Perl[89] suggested substituting differentiation by plotting $f(t+\Delta t)$ against $f(t)$ with constant Δt throughout. The main disadvantage of these two methods is that they emphasize high-frequency noise. Better results may be obtained by employing advanced differentiation algorithms. For example, one can consider the differentiation as an ill-posed problem[90,91] and apply Tikhonov regularization (see Sec. VI C 4), or make a spline interpolation of the data and then differentiate the result.[92,93]

Finally, Isenberg *et al.*[94] and Kirchner *et al.*[82] proposed using exponential depression to suppress the baseline. In this method, the input transient is multiplied by a decaying exponential $\exp(-\lambda_d t)$. The extracted amplitudes are unchanged, and the emission rates λ_i are related to the extracted λ_i through a simple relation $\lambda_i=\lambda_i-\lambda_d$. However, this treatment was regarded by Gardner *et al.*[95] as ''a drastic and unrecommended step.''

C. Smoothing of transients

The amplitude of high-frequency noise in the transient, which makes the problem of exponential analysis less stable, can be decreased by smoothing the transients. It is important that the smoothing algorithm should not introduce distortions into the component time constants. There are essentially two methods which were developed for the smoothing of exponential decays.

The method developed by Dyson and Isenberg,[96] also known as mean displaced ratio (MDR) method, is based on the assumption that the observed decay is a sum of an unknown number of discrete exponentials. The original function $f(t)$ is replaced by a new function $Y(q)$ that has the same time constants as the original curve, but a much improved SNR. For continuous data, $Y(q)$ is defined as

$$Y(q)=\frac{1}{T-Q}\int_0^{T-Q}[f(t+q)/f(t)]dt, \tag{19}$$

where $0\leqslant q<Q$ and $Q<T$, where T is the upper limit of t and Q is a positive constant that determines the degree of smoothing. Note that while the original observations are collected between $t=0$ and $t=T$, the smoothed data cover a somewhat smaller range, from $q=0$ to $q=Q$.

When the experimental situation yields $N+1$ discrete observations, made at equal increments of time Δt so that $t_i=i\Delta t$, the mean displaced ratio is defined as

$$Y_k=\frac{1}{N-L+1}\sum_{i=0}^{N-L}\frac{f_{i+k}}{f_i}, \tag{20}$$

where $k=0,1,\ldots,L$. The degree of smoothing is controlled by the choice of L, with increased smoothing achieved by decreasing the L value. Ikossi-Anastasiou *et al.*[83] and Dyson *et al.*[96] found that for the data with 5%–10% noise, $L=0.9N$ is a good choice.

If the original observations can be represented as

$$f(t)=\sum_{j=1}^{n} A_j \exp(-t/\tau_j), \tag{21}$$

then, as it is easy to obtain from Eqs. (20) and (21), the output curve has the form

$$Y(q)=\sum_{j=1}^{n} \beta_j \exp(-q/\tau_j), \tag{22}$$

where

$$\beta_i=\frac{A_j}{N-L+1}\sum_{j=1}^{n} \exp(-t_i/\tau_j)/f_i \tag{23}$$

and

$$Y_0=\sum_{i=1}^{n} \beta_i=1. \tag{24}$$

Dyson *et al.*[96] noted that the advantages of the MDR are twofold: not only does it reduce random deviations from the true decay without the risk of introducing systematic errors, but it also eliminates one unknown by exactly defining the sum of the amplitudes β_i [Eq. (24)]. It was also noted[83] that it is important to extract the baseline offset (Sec. V B) prior to implementing the MDR smoothing.

Provencher[97,98] suggested another smoothing algorithm:

$$Y(t_k)=\sum_{m=1}^{L} f(t_m)f(t_m+t_k-t_1) \Big/ \sum_{m=1}^{L} f^2(t_m), \tag{25}$$

where $k=1,2,\ldots,j$, $j=N-L+1$, and N is the number of data points. The output signal is given by Eq. (22), where the coefficients β_i are expressed as

$$\beta_j=A_j\sum_{m=1}^{L} f(t_m)\exp(-\lambda_j t_m) \Big/ \sum_{m=1}^{L} f^2(t_m). \tag{26}$$

However, since this algorithm uses a product rather than a ratio of the data points, it does not uniquely define the sum of the amplitudes, and therefore does not reduce the number of unknown parameters as the mean displaced ratio does.

Windowing filtering techniques, which are based on a convolution of the transient with a function with a band-limited Fourier image[99,100] can affect the exponential nature of the data, unless the signal is filtered after it is reversed in time.[101] The problem of detection and removal of impulse distortions of the experimental data was intensively discussed in the literature on digital image processing (see, e.g., Refs. 102–105).

VI. NUMERICAL PROCEDURES FOR EVALUATION OF TRANSIENTS

The first part of this review article discussed the question, ''why it is hard to analyze the transients,'' ''how the resolution of the exponential analysis can be improved by using the correct procedure of data acquisition.'' Yet, so far we have not discussed actual algorithms for extracting decay rates and amplitudes from the transients. In this section, we present a summary of major numerical algorithms used for exponential analysis in applied physics. This section is intended for physicists who wish to gain a deeper understanding of the techniques of exponential analysis. First-time readers who are interested in general features of exponential analysis rather than in details of the algorithms are advised to skip this section and continue reading from Sec. VII. The goal of this section is to classify the methods of exponential analysis, discuss their essential features, and give some guidelines to choose the most suitable for a given task algorithm. This review article is not intended to be a complete guide for programming of the discussed methods, and does not necessarily describe all mathematical aspects of the algorithms. This particularly applies to the most sophisticated methods, discussed in Sec. VI C. The reader interested in the details of correspondent algorithms should refer to the original articles cited in the text. Furthermore, we want to emphasize (and will return to this point later in the discussion) that the program code for a number of most of the complicated algorithms discussed in Sec. VI C is readily available from program libraries, and we strongly encourage readers to use the existing code rather than to write their own.

The structure of this section is as follows. Following the classification introduced in Sec. II, we divided the methods for exponential analysis into three large groups: monoexponential analyses based on the assumption that the decay consists of a single exponential; multiexponential analyses, which assume that decays consist of a sum of several (from 2 to 4) exponentials; and spectroscopic methods of exponential analysis, which do not make any assumptions about the decay rate spectrum and can be applied to both discrete exponentials and continuous distributions of emission rates.

The algorithms for the solution of the Laplace integral equation are derived assuming that both the exponential decay $f(t)$ and the spectral function $g(\lambda)$ are analytical functions. Some authors prefer to emphasize the discrete nature of the problem and formulate the Laplace equation in the matrix form $\mathbf{f}=\mathbf{Kg}$. In this equation $\mathbf{f}$ and $\mathbf{g}$ are data vectors and $\mathbf{K}$ is the matrix, which represents the Laplace integral operator. The algorithms are essentially the same for both matrix and analytical representations. Overviews and comparative analysis of major matrix-based methods can be found in Hansen[106–108] or Varah.[109] The matrix methods are often faster, at least if the matrices are not too large. Moreover, the matrix $\mathbf{K}^{-1}$ has to be regularized and calculated only once and can thereafter be repeatedly used for evaluation of different transients with the same SNR. The analytical methods do not have this advantage, but they work equally well with both small and very large amounts of data and, from our point of view, are easier to explain. Therefore,

our discussion will be confined to the analytical versions of the algorithms.

To avoid possible confusion, we want to emphasize that the methods are grouped according to the capabilities of the methods rather than according to the type of the transient. The methods of monoexponential analysis discussed in Sec. VI A can be applied to only monoexponential transients. However, *monoexponential transients* can be evaluated using any one of the discussed methods, including those from Secs. VI B and VI C. If the task of the reader is to analyze a monoexponential transient, we would also strongly recommend reading Sec. VI B on methods for multiexponential analysis.

A. Methods for monoexponential analysis

1. Fourier transform of the transients

The Fourier transform maps the exponential relaxation into a function of frequency according to the equation $F(\omega)=\Im[A\exp(-\lambda_0 t)+B]$, where $\Im$ is the Fourier transform operator. A very useful property of the Fourier transform of an exponential decay is that for any nonzero angular frequency ω, the ratio of the real and the imaginary components of $F(\omega)$ yields the quantity $-\lambda_0/\omega$ independently of the baseline offset B or amplitude A (Ref. 82):

$$\lambda_0=-\omega\,\mathrm{Re}[F(\omega)]/\mathrm{Im}[F(\omega)]. \tag{27}$$

Although the decay constant λ_0 can be determined from Eq. (27) for any frequency ω, usually only the lowest frequencies are used.[110–113] After λ_0 is calculated, the amplitude of the decay A and the baseline B can be easily calculated:

$$A=\frac{(\omega^2+\lambda_0^2)}{\lambda_0}\times\frac{\mathrm{Re}[F(\omega)]\Delta t}{1-\exp(-\lambda_0 t_m)}, \tag{28}$$

$$B=F(0)/N-(A_0/\lambda_0 t_m)[1-\exp(-\lambda_0 t_m)], \tag{29}$$

where ω is any nonzero Fourier component given by $2\pi n/t_m$, and $\Delta t=t_m/N$. Here, t_m is the time of the last observation, and N is the number of data points in Fourier transform. An important feature of this method is that it is noniterative, and hence is very fast. The method gives good results even if the noise level is as high as 10%.[82,83] The important advantage of this method is that it can be realized on a computer (using the fast Fourier transform algorithm) or using electronic spectrum analyzers. Another advantage of the method is that it determines the baseline offset (which as we will show below only few methods can do) and thus does not require that the baseline be subtracted before the method is applied to a transient.

2. Algebraic methods

Devries and Khan[114,115] proposed the ''divisor method'' based on the formula of Moore *et al.*[84] Assuming that the baseline offset is extracted (Sec. V B), it is easy to obtain from the expression $f(t_i)=A\exp(-t_i/\tau)$ for $i=1,2$ the value of the decay time constant

$$\tau=(t_2-t_1)/\ln[f(t_1)/f(t_2)]. \tag{30}$$

The necessity of subtracting of the offset can be easily overcome by measuring three values of $f(t)$ instead of two:[116,117]

$$\frac{1}{\tau}=\frac{1}{\Delta t}\ln\left(\frac{f(t_2)-f(t_1)}{f(t_3)-f(t_2)}\right), \tag{31}$$

where $\Delta t=t_2-t_1=t_3-t_2$. Mukoyama[118] and later Zubkov *et al.*[119] and Kim *et al.*[120] proposed a four-point formula:

$$\tau=(t_3-t_1)\times\ln\left(\frac{C(t_2)-C(t_1)}{C(t_4)-C(t_3)}\right), \tag{32}$$

where $t_2-t_1=t_4-t_3$. Note that Eq. (32) reduces to Eq. (31), if $t_2=t_3$. A disadvantage of these algebraic methods as compared to the Fourier transform method (Sec. VI A 1) is that it uses only 2 to 4 data points to compute the time constants, and thus is more sensitive to deviations of data points due to noise. Another drawback of the methods from this subsection is that they cannot be used to determine the amplitude of the decay and are not tolerant of nonzero baseline offsets.

B. Methods for multiexponential analysis

1. Graphical analysis technique (peeling method)

The graphical analysis technique (also known as the peeling method) was widely used in the 1960s, when desktop computers were not available. If the components λ_i of the multiexponential decay are not too close to each other, then the transient $f(t)$ for large values of t can be approximated by the slowest component, i.e.,

$$\ln[f(t)]=\ln\left(\sum_{i=1}^{n}A_i\exp(-\lambda_i t)\right)$$
$$\approx\ln[A_n\exp(\lambda_n t)]\equiv\ln(A_n)-\lambda_n t. \tag{33}$$

Obviously, A_n and λ_n can be determined, plotting the transient $f(t)$ in a semilog scale and fitting its tail with a straight line. Then, the slowest component $A_n\exp(-\lambda_n t)$ is subtracted from the transient $f(t)$ and the procedure is repeated to determine the parameters of the second slowest component, etc. According to Van Liew,[121,122] up to three exponentials can be extracted by this method. The peeling method requires that the baseline offset be extracted before Eq. (33) is applied to the transient.

With the development of computers, graphical analysis methods transformed into nonlinear least squares fitting routines[123] discussed in Sec. VI B 2.

2. Nonlinear least squares analysis

The nonlinear least squares (NLS) analysis consists in minimization of the functional

$$\chi^2(\mathbf{p})=\sum_{j=1}^{N}[f_{\exp}(t_j)-f_0(t_j,\mathbf{p})]^2,$$
$$f_0(t,\mathbf{p})=\sum_{i=1}^{n}A_i\exp(-\lambda_i t)+B \tag{34}$$

by variation of $J=2n+1$ parameters p_i, the parameters being amplitudes A_i, decay rates λ_i, and eventually an offset B, where n is the assumed number of exponential components, and N is the number of experimental points. Minimization of the value of χ^2 is achieved by variation of the parameters p_i. Sometimes, the right-hand side of Eq. (34) is

multiplied by weight factors. These weight factors enable one to emphasize or deemphasize certain data points. However, we will not discuss these weight factors since, to the best of our knowledge, they are seldom used in exponential analysis.

The search for the solution is an iterative procedure. Each iteration consists of two parts. First, a decision is made in which direction to proceed in the parameter space, and then it is calculated how far to go in that direction. Modifications of the nonlinear least squares method differ mainly in procedures used to determine the direction and the length of the steps, and can be classified according to the maximum order of derivative used in calculations.

The simplest fitting routines do not use derivatives at all. For example, the grid search minimizes χ^2 with respect to each parameter separately, varying one parameter at a time. Each step in this minimization procedure is a single parameter search, and a quadratic approximation is used to find the minimum (see Bevington[124]). The efficiency of a grid search can be improved, using conjugate directions methods.[125–127] Another nonderivative method is the simplex method.[100,128–130]

Slope-following methods (also known as gradient search methods) evaluate the first derivatives of the error function $\partial\chi^2/\partial p_j$ in order to determine how the parameters should be changed to minimize the χ^2. All parameters p_i are incremented simultaneously, moving the parameter vector a short distance $\delta\mathbf{p}$ in the direction opposite to that of the gradient vector $\nabla\mathbf{p}$:

$$\nabla\mathbf{p}=\sum_{j=1}^{J}\left(\frac{\partial\chi^2(\mathbf{p})}{\partial p_j}\hat{p}_j\right), \tag{35}$$

where p_j are the parameters and $\hat{p}_j$ indicate unit vectors. Equation (35) provides only the direction and not the length of the parameter increment $\delta\mathbf{p}$. The latter is obtained by a one-dimensional search. This procedure is repeated until χ^2 converges to a minimum. The other modification of the gradient search method is the method of conjugate gradients.[131–133] This method determines the search direction on each step as a linear combination of gradients on the current and preceding steps.

Methods that use the second-order derivatives calculate both the direction and the size of a step. Methods that use the second-order derivatives calculate both the direction and the size of a step. To approximate the dependence $\chi^2(\mathbf{p})$ in Eq. (34) in the Gauss–Newton method the fitting function is expanded to the first order in a Taylor series:

$$f(t,\mathbf{p}+\delta\mathbf{p})\approx f_0(t,\mathbf{p})+\sum_{j=1}^{J}\left(\frac{\partial f_0(t,\mathbf{p})}{\partial p_j}\delta p_j\right), \tag{36}$$

where $f_0(t,\mathbf{p})$ is the model function Eq. (34) before the step, and $f(t,\mathbf{p}+\delta\mathbf{p})$ is the function after the step. The Newton–Raphson method uses the Taylors expansion of χ^2:

$$\chi^2(\mathbf{p}+\delta\mathbf{p})\approx\chi_0^2(\mathbf{p})+\sum_{j=1}^{J}\left(\frac{\partial\chi_0^2(\mathbf{p})}{\partial p_j}\delta p_j\right), \tag{37}$$

where $\chi_0^2(\mathbf{p})$ and $\chi_0^2(\mathbf{p}+\delta p)$ are the values of χ^2 before and after the step.

The optimum values for the parameter increments δp_j are those for which the function χ^2 is a minimum in the parameter space, i.e., for which the derivatives of χ^2 with respect to the parameters are zero. This condition together with Eqs. (36) and (37) results in a set of J simultaneous linear equations in δp_j, which can be presented as a matrix equation (see, e.g., Ref. 124):

$$\beta_k=\sum_{j=1}^{J}\delta p_j\alpha_{jk},\quad k=1,\ldots,N, \tag{38}$$

the curvature matrix α and the gradient β are given for the Newton–Raphson method by

$$\beta_k=-\frac{1}{2}\frac{\partial\chi_0^2(\mathbf{p})}{\partial p_k},\quad \alpha_{jk}=\frac{1}{2}\frac{\partial^2\chi_0^2(\mathbf{p})}{\partial p_j\partial p_k}, \tag{39}$$

and for the Gauss–Newton method by

$$\beta_k=-\frac{1}{2}\frac{\partial\chi_0^2(\mathbf{p})}{\partial p_k}=\sum_{i=1}^{N}[f_{\exp}(t_i)-f_0(t_i,\mathbf{p})]\frac{\partial f_0(t_i,\mathbf{p})}{\partial p_k},$$

$$\alpha_{jk}=\sum_{i=1}^{N}\frac{1}{2}\frac{df_0(t_i,\mathbf{p})}{\partial p_j}\frac{\partial f_0(t_i,\mathbf{p})}{\partial p_k}. \tag{40}$$

The Gauss–Newton and Newton–Raphson methods differ mainly in the formulas for calculation of the curvature matrix α. The Gauss–Newton method is used more often because it estimates the second derivatives from the first derivatives instead of calculating them directly. This greatly simplifies calculations. Several modifications of the Gauss–Newton method were reported that avoid calculations of the derivative at each step by approximating them from the residuals. These methods are known as quasi-Newton methods.[134–139] As a rule, the Gauss–Newton or Newton–Raphson methods and their modifications require more computations for one step, but converge in a fewer number of steps than gradient or grid search methods and, as a consequence, have better overall performance. This was shown in numerous tests.[140–147]

The direction and length of each iteration are determined by the inverse matrix α^{-1}. It must be a positive definite matrix for the iterative process to proceed toward the minimum. Therefore, Gauss–Newton and Newton–Raphson methods can only be used in a close vicinity of a minimum. Crockett *et al.*[148] and Greenstadt[149] have shown that the ratio of the largest to the smallest eigenvalue of α^{-1} provides a measure of the convergence to a solution. (See also the review of Spang.[150])

If the starting point of the search is far from the minimum and the approximation [Eqs. (36) and (37)] fails, the search procedure tends to overshoot the minimum and diverge. In this case, one can use partial steps $k\Delta\mathbf{p}$, with $k<1$, instead of the full step $\Delta\mathbf{p}$ specified by the Gauss–Newton or Newton–Raphson algorithms. This method[140,151] is known as ''damped least squares.'' The optimum value of the ''damping coefficient'' k can be obtained by using a one-dimensional search procedure.[152]

Another modification of the Gauss–Newton method was proposed by Levenberg[153] and, independently, by Marquard.[144] It consists of replacing the curvature matrix α by $\alpha+\gamma I$, where I is an identity matrix. This method may be viewed as a compromise between the Gauss–Newton and steepest descent methods. When $\gamma=0$, it gives a Gauss–Newton step. As $\gamma\to\infty$, the step direction coincides with the direction of the gradient, and the value of γ serves as the ''damping'' coefficient. The usual practice is to adjust the value of γ dynamically from one iteration to the next by a rule which is based on the past behavior of the algorithm. An analytical approach to the choice of γ was proposed by Goldfeld *et al.*[154]

An important advantage of the Levenberg–Marquard method is that even if the matrix α is singular, the matrix $\alpha+\gamma I$ is nonsingular whenever $\gamma>0$ and can be made well conditioned by choosing γ to be sufficiently large. This is a very useful property for the fitting of exponentials because if the time constants of two decays become close, the matrix α becomes nearly singular.

A comprehensive treatment of the Gauss–Newton, Newton–Raphson and Levenberg–Marquard algorithms and listings of computer programs can be found in Bevington[124] and in Press *et al.*[100] Worsley and Lax[155] and Flanagan *et al.*[152] applied the Gauss–Newton method to a linear combination of exponentials and analyzed its accuracy. The determination of the standard deviation of parameters in the nonlinear least squares method was discussed by Burrell[156] and Swarte.[157] Further modifications to improve convergence of the Gauss-like methods were proposed by Deuflhard *et al.*,[158] Neri *et al.*,[159] Lybanon,[160,161] Dennis *et al.*,[145] Hartley,[162] and Lee *et al.*[163]

The iterative least squares analysis is a simple, easy-to-implement and efficient technique, and is widely used for exponential analysis. In the case of a monoexponential transient, the least squares algorithm can even be reduced to an algebraic equation that can be solved directly.[164] However, two essential difficulties are inherent in the least squares fit: the initial guesses for the values of the unknown decay parameters are needed and, if these initial estimates are poor, the iteration may converge to a local minimum rather than to the absolute minimum.[165–167]

The nonlinear least squares (NLS) method was successfully used for analysis of decays that consisted of one to three exponential decays. Grinvald,[167] Morimoto *et al.*,[168,169] and Tahira *et al.*[170] found that for a SNR better than 10^3 a double-exponential decay could be resolved if the ratio of time constants was $\tau_1/\tau_2>2$. However, in the case of three exponents the method was found to be very sensitive to even a small amount of noise.[171] Bromage[172] concluded that for a decay-time-constant ratio of 2:1, the retrieved values of parameters are about 200 times more uncertain on going from two to three exponentials. Clayden *et al.*[173] concluded that in the case of a double-exponential decay, NLS fit is robust if $\tau_1/\tau_2>2.4$ and SNR is about 10^3. If a decay consists of three exponential and SNR$\approx 10^3$, the components can be resolved if $\tau_1/\tau_2=\tau_2/\tau_3>3.5$. If the SNR is about 100, three components can be resolved if the decay-time-constant ratio is above 10. It was also noted[174,175] that the least squares method is not robust with respect to nonrandom errors in the transient.

3. Algebraic methods (Prony's method)

Prony's method requires that the decay $f(t)$ be measured at $2n$ equidistant points $f(t_0)$, $f(t_1)$, ..., $f(t_{2n-1})$. Prony[29] showed that approximation of the decay by a sum of n exponentials,

$$f(t)=\sum_{i=1}^{n} A_i \exp(-\lambda_i t)=\sum_{i=1}^{n} A_i \mu_i^t, \quad \mu_i=\exp(-\lambda_i) \tag{41}$$

results in a system of $2n$ equations with $2n$ unknowns: $A_1, A_2, \ldots, A_n, \mu_1, \mu_2, \ldots, \mu_n$:

$$\begin{cases} A_1+A_2+\cdots+A_n=f(t_0) \\ A_1\mu_1+A_2\mu_2+\cdots+A_n\mu_n=f(t_1) \\ A_1\mu_1^2+A_2\mu_2^2+\cdots+A_n\mu_n^2=f(t_2) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ A_1\mu_1^{2n-1}+A_2\mu_2^{n-1}+\ldots+A_n\mu_n^{2n-1}=f(t_{2n-1}) \end{cases} . \tag{42}$$

Solving this equation, one finds the decay rates λ_i and amplitudes A_i. The difficulty with the solution of equation Eq. (42) is that they are nonlinear in the μs. This difficulty can be minimized by a method proposed by Prony[29] in 1795. (A description of this method can be found in Refs. 176–179.) He suggested to concentrate the nonlinearity of the system in a single algebraic equation. Let $\mu_1, \mu_2, \ldots, \mu_n$ be the roots of the equation

$$(\mu-\mu_1)(\mu-\mu_2)\cdots(\mu-\mu_n)=0, \tag{43}$$

that can be rewritten as

$$\mu^n-\alpha_1\mu^{n-1}-\alpha_2\mu^{n-2}-\cdots-\alpha_{n-1}\mu-\alpha_n=0. \tag{44}$$

In order to define the coefficients α, we multiply the first equation in Eq. (42) by α_n, the second equation by $\alpha_{n-1}, \ldots$, the nth equation by α_1 and the $(n+1)$th equation by -1. If we then add up the first n equations and use Eq. (44), we obtain

$$f(t_n)-\alpha_1 f(t_{n-1})-\ldots-\alpha_n f(t_0)=0. \tag{45}$$

Starting successively with the second, third, ..., nth equations, we find that Eqs. (42) and (44) imply the following n linear equations:

$$\begin{cases} f(t_{n-1})\alpha_1+f(t_{n-2})\alpha_2+\ldots+f(t_0)\alpha_n=f(t_n) \\ f(t_n)\alpha_1+f(t_{n-1})\alpha_2+\ldots+f(t_1)\alpha_n=f(t_{n+1}) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ f(t_{2n-2})\alpha_1+f(t_{2n-3})\alpha_2+\ldots+f(t_{n-1})\alpha_n=f(t_{2n-1}) \end{cases} . \tag{46}$$

We may therefore first solve the linear equations [Eq. (46)] for $\alpha_1, \alpha_2, \ldots, \alpha_n$, then find the roots $\mu_1, \mu_2, \ldots, \mu_n$ of Eq. (44), and finally solve any n of the linear equations [Eq. (42)] for $A_1, A_2, \ldots, A_n$.

This method is also known as the ''operator method'' (Ref. 180) or the method of ''linear predictive modeling'' (Ref. 70). This term reflects that a linear combination of values $f(t)$ in the ''past'' is related to the value in the ''fu-

ture'' by the linear operator equation [Eq. (46)]. In the covariance method of linear predictive modeling, applied to the problem of transient analysis by Shapiro *et al.*,[70] Enderlein *et al.*[181] and Apanasovich *et al.*,[182] equations equivalent to Eqs. (44) and (46) were obtained from quite different considerations, namely from application of the z transform[183] to Eq. (41).

The method of Prony in its original form is extremely sensitive to noise and round-off errors[184] and is not acceptable for analysis of real (noisy) transients. Hildebrand[178] proposed using more than $2n-1$ experimental noisy points $f(t_k)$ for the determination of $2n$ parameters and search for $\alpha_1, \alpha_2, \ldots, \alpha_n$ as a linear least-square solution of Eq. (46), i.e., as a minimum of the functional

$$\chi^2 = \sum_{k=n+1}^{M} \left(f(k) - \sum_{m=1}^{n} \alpha_m f(km) \right)^2, \tag{47}$$

where M is the number of measured data points, $M > 2n-1$. Another version of the Prony method which utilizes more that $2n-1$ data points was developed by Sun *et al.*,[86,185] who presented Eq. (46) in the matrix form $\mathbf{Y} = \mathbf{AX}$, where

$$\mathbf{Y} = \begin{bmatrix} f(t_n) \\ f(t_{n+1}) \\ \cdots \\ f(t_{2n-1}) \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} f(t_{n-1}) & \cdots & f(t_0) \\ f(t_n) & \cdots & f(t_1) \\ \cdots & \cdots & \cdots \\ f(t_{2n-2}) & \cdots & f(t_{n-1}) \end{bmatrix},$$
$$\mathbf{A} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \cdots \\ \alpha_n \end{bmatrix}. \tag{48}$$

Instead of the straightforward solution of this matrix equation in the form $\mathbf{A} = [\mathbf{X^T X}]^{-1} \mathbf{X^T Y}$, they introduced an auxiliary matrix $\mathbf{Z}$ determined as

$$\mathbf{Z} = \begin{bmatrix} f(t_{n-1+k}) & \cdots & f(t_k) \\ f(t_{n+k}) & \cdots & f(t_{k+1}) \\ \cdots & \cdots & \cdots \\ f(t_{2n-2+k}) & \cdots & f(t_{n-1+k}) \end{bmatrix}, \tag{49}$$

where k is an integer $k \geqslant 1$, and the total number of points in the decay is $M = 2n-1+k$. Sun *et al.*[86,185] proposed to compute the solution as $\mathbf{A} = [\mathbf{Z^T X}]^{-1} [\mathbf{Z^T Y}]$. Such a solution establishes a correlation between the values of function f in the future and in the past and thus decreases the sensitivity of the solution to noise. In this form, Prony's method was applied to compute solutions for double[86] and triple-exponential decays[185] and showed a comparable accuracy with that of the NLS method (Sec. VI B 2). However, Prony's method is much more computationally effective. On the other hand, the baseline offset should be removed from the transient before Prony's method is applied, which is not necessary for NLS.

A very important problem for Prony's method is to determine the number of exponential components in the transient n, i.e., the rank of the matrix $\mathbf{A}$. Kumaresan *et al.*[186] suggested using Prony's method first with n larger than the number of exponential components L, which is actually expected. The result is a set of n exponentials that are candidates for the signal components. The next step is then to determine a smaller subset of size L out of n exponentials for which a linear combination of L exponentials best approximates the observed data using the least squares criterion. They suggested starting with a subset of the size $L=1$ and then increase it by one until no substantial improvement in the error is observed. Holt *et al.*[187] suggested singular value decomposition of the matrix $\mathbf{X}$ [Eq. (48)] to determine the number of exponential components. They suggested increasing the size of $\mathbf{X}$ (i.e., the number of unknown components α) as long as all singular values of $\mathbf{X}$ remain large. By successively increasing the size of $\mathbf{X}$, a system must be eventually reached which is unacceptably ill conditioned. Then, some of the singular values of $\mathbf{X}$ will become very small. The transition from a well-conditioned system to such an ill-conditioned system is expected to occur when the size of the matrix $\mathbf{X}$ exceeds the number of exponential components.

4. Differentiation of transients and method of modulating functions

As with Prony's method, the differentiation method is based on a certain equation which an exponential decay of the form Eq. (4) should comply with. It can be easily shown that a transient given by Eq. (4) obeys the following differential equation:

$$f(t) + \alpha_1 \frac{df(t)}{dt} + \alpha_2 \frac{d^2 f(t)}{dt^2} + \cdots + \alpha_n \frac{d^n f(t)}{dt^n} = 0, \tag{50}$$

where

$$\alpha_1 = \sum_i \lambda_i^{-1}, \quad \alpha_2 = \sum_{i<j} (\lambda_i \lambda_j)^{-1}, \ldots,$$
$$\alpha_n = (\lambda_1 \lambda_2 \cdots \lambda_n)^{-1}. \tag{51}$$

In the case of a noise-free decay $f(t)$, the parameters $\alpha_0, \alpha_1, \ldots, \alpha_n$ can be determined by calculating derivatives of $f(t)$ and solving Eq. (50) for $t = t_1, t_2, \ldots, t_n$. For a monoexponential decay, the solution of Eqs. (50) and (51) is simply

$$\lambda = -\frac{df(t)}{dt} \Big/ f(t). \tag{52}$$

A disadvantage of the differentiation method is that it does not allow one to determine the amplitudes of decay. However, the major problem with this algorithm is that experimental transients are noisy, and large errors would result from their numerical differentiation.

Differentiation of noisy transients can be avoided by using the method of modulating functions proposed by Loeb and Cahen.[188,189] Modulating functions $\varphi(t)$ are functions that satisfy the condition that the function and all its derivatives are equal to zero for $t=0$ and $t=T$, assuming that the function $f(t)$ is determined on the interval $t=[0,T]$. Examples of such functions are $t^{n+1}(T-t)^{n+1}$, $\sin[(\pi/T)(t-T)]$, etc. It can be easily shown that

$$\int_0^T \frac{d^n f(t)}{dt^n}\varphi(t)dt$$
$$= \frac{d^{n-1}f(t)}{dt^{n-1}}\varphi(t)\Big|_0^T - \int_0^T \frac{d^{n-1}f(t)}{dt^{n-1}}\frac{d\varphi(t)}{dt}dt$$
$$= \ldots = (-1)^n \int_0^T f(t)\frac{d^n\varphi}{dt^n}dt. \tag{53}$$

Following the idea of Valeur and Moirez[190] and multiplying both parts of Eq. (50) by a modulating function $\varphi(t)$ and integrating from 0 to T by parts, as in Eq. (53), we obtain

$$\alpha_0 \int_0^T f\varphi\, dt - \alpha_1 \int_0^T f\frac{d\varphi(t)}{dt}dt + \cdots + (-1)^n\alpha_n \times \int_0^T f\frac{d^n\varphi}{dt^n}dt = 0. \tag{54}$$

Hence, the derivatives of $f(t)$ in Eq. (50) are substituted by the derivatives of $\varphi(t)$, which can be calculated analytically. It should be noted that one needs a system of n equations to determine n unknown parameters α_n. In Eq. (50) such a system could be obtained by substituting the function $f(t)$ by its values at different times $t=t_1,t_2,\ldots,t_n$. This cannot be done with Eq. (54) since it includes integration over t. However, one can use n different modulating functions, for example of the form $\varphi(t)=t^{n+1}(T-t)^{n+1+P}$, where $P=1,2,\ldots,n$.

To the best of our knowledge, the sensitivity of the method of modulating functions to noise has never been analyzed using simulated decays. However, it was reported that the method was applied to pulse fluorometry[191] and capacitance spectroscopy[192] and could successfully resolve two experimentally measured exponentials with a ratio of time constants of 2. The method of modulation function does not enable one to determine the amplitude of the transients and is not tolerant to nonzero baseline offsets.

5. Integration method

The integration method was proposed by Tittelbach-Helmrich.[193] The idea of this method is somewhat similar to the differentiation method, discussed in the previous section. However, instead of differentiation, the exponential decay $f(t)$ [given by Eq. (4)] is integrated over time $p=1,2,\ldots,n$ times. It can be then shown that the following equation holds:

$$f(t)+b_1f^{(1)}(t)+b_2f^{(2)}(t)+\cdots+b_Nf^{(n)}(t)=0, \tag{55}$$

where $f^{(p)}(t)$ are the functions obtained by a p-fold integration of $f(t)$.

$$f^{(p)}(t)=\int_{(p)}\cdots\int_t^\infty f(\theta)d\theta=\sum_{k=1}^n A_k\tau_k^p\exp(-t/\tau_k) \tag{56}$$

and the coefficients b_i are given by

$$b_1=-\sum_i \lambda_i, \quad b_2=\sum_{i<j}\lambda_i\lambda_j,\ldots,$$
$$b_N=(-1)^n\lambda_1\lambda_2\cdots\lambda_N. \tag{57}$$

For example, for a monoexponential transient, it is easy to show that Eqs. (55) and (57) reduce to

$$\lambda=\frac{f^{(1)}(t)}{f(t)}. \tag{58}$$

Since integration in Eq. (56) is done on an interval $[t,\infty]$, the functions $f^{(p)}(t)$ remain functions of t, and the unknown coefficients b_p can then be obtained from Eq. (55) by solving a system of equations for any N data points $t=t_i$, or performing a least squares fit using all experimental data points. Having determined the B_p, the decay rates λ_i can be determined by solving the system of equations, Eq. (57).

Tittelbach-Helmrich[193] showed that when integration is performed over a limited time interval, the integration algorithm also enables one to determine the baseline. However, the solution is more stable and exact[193] if the baseline is subtracted from the raw signal before the analysis. For the integration method, it was demonstrated[193] on computed two-component decays with $\tau_1/\tau_2=2.5$ that reliable separation of the components is impossible if the SNR is less than 30. For $\tau_1/\tau_2=5$, two components can be separated up to a SNR of about 10. When two-component analysis is performed with a baseline determination, the mean error of the calculated time constants increased by a factor of about 3–10 compared to the case without baseline restoration.

6. Method of moments

The theory of moments itself dates back to the nineteenth century, although it was applied to the exponential analysis much later by Bay.[194] This method is based on the evaluation of the time-weighted moments μ_k of the transient $f(t)$ defined by

$$\mu_k=\int_0^\infty t^k f(t)dt, \tag{59}$$

where $k\geqslant 0$. For a monoexponential $f(t)=\exp(-t/\tau)$, the moments μ_k are smooth functions of decay time constant with a maximum at $t=\tau/k$ (Ref. 96). Defining the parameters G_s as

$$G_s=\sum_{i=1}^n A_i\tau_i^s, \tag{60}$$

where $s\geqslant 1$, it can be shown[83] that the G_s are related to the moments of the decay $f(t)$ by the equations:

$$G_s=\mu_{s-1}/(s-1)! \tag{61}$$

Equation Eq. (61) can be used to determine G_s. Then, the individual time constants can be determined by solving a system of equations:[195]

$$\begin{vmatrix} 1 & \tau & \dots & \tau^n \\ G_1 & G_2 & \dots & G_{n+1} \\ \dots & \dots & \dots & \dots \\ G_n & G_{n+1} & \dots & G_{2n} \end{vmatrix} = 0, \tag{62}$$

where the roots of the determinant are the individual time constants $\tau_1, \tau_2, \dots, \tau_n$ of the exponential components. After the time constants are determined, the amplitudes A_i are found by using Eq. (60). The method of moments can be applied to analysis of decays that originate from a nonabrupt excitation function, which is the case in, e.g., fluorescence decay studies (see Refs. 83, 94, 96, and 195–200 for more detail).

An important part of the analysis of decays using method of moments is the determination of the number of exponential components. A common method, discussed by Isenberg *et al.*[94,96] and Kirchner *et al.*,[82] is to fit the data successively to an increasing number of exponential components. As the number of components is incremented, each succeeding fit incorporates the components of the preceding fit. The appropriate number of components is finally determined by evaluating the rms error and by examining the relevance of the results. It is expected that for an n component transient, analysis for $n+1$ components will result in an additional, superfluous component with either an exceedingly small or negative amplitude, or a negative time constant. Finally, before one can apply the method of moments, the baseline offset must be extracted from the measured data. Without this step, an excessively large contribution is introduced into the moments integrals from the tail of the transient.

The method of moments works very well with monoexponential decays and returns reliable results even in the case of SNR as low as 10 (Ref. 82). The method of moments also proved to be capable of extracting two components, even in relatively severe noise (SNR=10).[82] Isenberg[171] showed that the method of moments can successfully resolve three exponential components with $\tau_1/\tau_2/\tau_3 = 3:7:11$, and even concluded that the method of moments was more stable with respect to noise than NLS.

7. Rational functions (Laplace-Padé) approximation

The method of rational approximation for an approximate Laplace transform inversion was suggested by Luke,[201] Fair[202] and developed by Longman[203–206] and Akin and Counts.[207,208] The idea of the rational function approximation method is to approximate $f(t)$ by a convergent series of rational functions $\{f_k(t)\}$. This series is then inverted analytically to give a sequence of functions $\{g_k(\lambda)\} = \{L[f_k(t)]\}$ which, if properly chosen, will converge rapidly to $g(\lambda)$ as $k \to \infty$. However, this method is used less often than the elegant rational functions approximation suggested by Yeramian *et al.*[209] and Aubard *et al.*[210] They proposed using the rational functions to approximate the Laplace image of the transient $f(t)$ rather than the transient itself. The Laplace transform, when applied to the function of the form

$$f(t) = \sum_{j=1}^{n} A_j \exp(-\lambda_j t) \tag{63}$$

gives the expression

$$L[f](p) = \int_0^\infty \exp(-pt) f(t)\, dt = \sum_{j=1}^{n} \frac{A_j}{p+\lambda_j}. \tag{64}$$

Detecting the exponential components is then a matter of identifying the poles of Eq. (64). This involves three steps. The first step is to express the Laplace transform of $f(t)$ at a specific point p_0 as a polynomial function through the use of a Taylor expansion truncated to the power K:

$$F(p) = L[f](p) \cong \sum_{j=0}^{K} c_j (p - p_0)^j, \tag{65}$$

$$c_j = \frac{1}{j!} \left(\frac{d^j L[f]}{dp^j} \right) (p_0), \tag{66}$$

where

$$\frac{d^j L[f]}{dp^j}(p) = \int_0^\infty (-t)^j f(t) \exp(pt)\, dt. \tag{67}$$

The second step is to describe the polynomial [Eq. (64)] in terms of its Padé approximant. The necessary algorithms for this step can be found in Perron[211] or Longman.[212] A Padé approximant, denoted by $[L/M](p)$, is the rational function obtained by the division of two polynomials $A_L(p)$ and $B_m(p)$ of degree L and M, respectively:

$$[L/M](p) = \frac{A_L(p)}{B_M(p)} = \frac{a_0 + a_1 p + \dots a_L p^L}{b_0 + b_1 p + .. b_M p^M}. \tag{68}$$

Decomposition of the Pade approximant into its partial fractions in the third stage gives the exponential time constants and amplitudes according to Eq. (64). It is assumed at this stage that the series represented by Eq. (64) is equivalent to the polynomial fraction, Eq. (68). Note that because of the form of the rational expression, Eq. (64), we are only interested in the $[n-1/n]$ approximants, where n is the expected number of exponential components in the multiexponential decay (usually 2 or 3).

The only input parameter of the method is the p_0 value. Although theoretically the solution should not depend on p_0, round-off errors may result in an unstable solution if the choice of p_0 is poor (see Ref. 210 for details). A potentially serious source of errors of the Laplace–Padé method is truncation of the measured decays.[213] Examples of applications of the Laplace–Padé technique can be found in Refs. 213, 214, 215. The Laplace–Padé approximation is not applicable to decays containing baseline offset. For the Laplace–Padé method it was reported that even with a SNR of about 15, component detection (for four-component decay with the ratio of the neighboring τ between 3 and 10) could be performed satisfactorily.[209] Clayden[214] showed that without noise one can resolve two exponentials with the ratio of time constants $\tau_1/\tau_2 > 1.5$ and three exponentials if $\tau_1/\tau_2 = \tau_2/\tau_3 > 2$.

C. Spectroscopic methods for analysis of nonexponential decays

1. Sampling methods

The sampling methods originate from ''delta convergence sequences'' (Ref. 216). They represent the spectral function $g(\lambda)$ as a linear combination of values of the data function $f(t)$ in a sequence of points $t_1, t_2, \ldots, t_n$ (Refs. 217–220).

As was shown by Davies *et al.*[221] and Nolte *et al.*,[222] the most accurate sampling formula was derived by Stehfest,[223] who based it on a statistical expectation function defined by Gaver.[224] Given a Laplace image, in our case an experimental decay curve $f(t)$, the algorithm calculates an approximation to the inverse $g(\lambda)$ as follows:

$$g(\lambda)=\frac{\ln(2)}{\lambda}\sum_{m=1}^{N} K_m f\left(\frac{m\ln 2}{\lambda}\right), \tag{69}$$

where

$$K_m=(-1)^{m+(N/2)} \times \sum_{k=(m+1)/2}^{\min(m,N/2)} \frac{(2k)!k^{1+(N/2)}}{(N/2-k)!k!(k-1)!(m-k)!(2k-m)!}. \tag{70}$$

Theoretically $g(\lambda)$ becomes more accurate with increasing N. However, rounding errors worsen the results if N becomes too large, because of numerous factorials in K_m. The optimum N is approximately equal[222,223] to the number of digits the computer is working with.

Another well-known formula is that of Post and Widder:[225,226]

$$g(\lambda)=\lim_{k\to\infty}\left[\frac{(-1)^k}{k!}\left(\frac{k}{\lambda}\right)^{k+1} f^{(k)}\left(\frac{k}{\lambda}\right)\right], \tag{71}$$

where $f^{(k)}$ denotes the kth derivative. It is clear that relatively small errors in the evaluation of the derivatives could seriously impair the accuracy of Eq. (71) since differentiation, in general, expands inaccuracies. Therefore, Eq. (71) is frequently used with only the first-order derivative:[217]

$$g(\lambda)=-\lambda^{-2} f^{(1)}(\lambda^{-1}) \tag{72}$$

or even with the zero-order derivative:

$$g(\lambda)=\lambda^{-1} f(\lambda^{-1}). \tag{73}$$

The last two formulas were successfully used in capacitance spectroscopy of semiconductors by Okushi and Tokumaru,[227–229] Tomokage *et al.*[230] and Ishikawa *et al.*[231] The sampling methods discussed in this section [with the exception of the method given by Eqs. (71) and (72)] require that the baseline offset be extracted before the method can be applied. The Gaver–Stehfest algorithm was tested by Nolte *et al.*[222] with double precision arithmetic and $N=24$. It was shown that without noise one can typically resolve two exponentials with a ratio of time constants of 1.5.

2. Correlation method

The correlation method is a signal processing method where the input signal is multiplied with a weighting function $W(t)$ defined on the time interval $[0,t_c]$, and the product is averaged over a period of correlation t_c. In general form, the output signal of a correlator is given by

$$S[g(\lambda),t_c,t_d]=t_c^{-1}\int_{t_d}^{t_d+t_c} f(t)W(t-t_d)dt, \tag{74}$$

where W is the weighting function and S is the output signal of the correlator, which is a function of the decay-rate distribution $g(\lambda)$, the duration of the weighting function t_c, and of the delay time t_d between the beginning of the decay and beginning of the correlation. The delay time is usually introduced to improve selectivity or to avoid distortions of the signal due to overload of the measurement system just after the excitation which triggers the decay.

The output signal of a correlator is in fact a Laplace image of the weighting function $W(t)$ [which is easy to show assuming $f(t)=\exp(-\lambda t)$]. Therefore, it is fairly simple and using any table of Laplace transformations, to find such a weighting function that the function $S(\tau_s)$ will give a maximum for a certain $\tau_s=\tau_{\max}$, and will drop to zero if $\tau_s\to 0$ or $\tau_s\to\infty$. The selectivity of a correlator and its sensitivity to noise in the input transients are strongly affected by the shape of the weighting function $W(t)$. Since correlation analysis of transients is a major technique for determination of decay time constants in DLTS,[34] much effort has been made to find a correlation function that would combine high resolution with high tolerance to noise in the transients. The results of these studies have been recently reviewed by Istratov *et al.* (see Ref. 232 and references therein). It has been shown[233,234] that a properly chosen correlation function can provide a resolution only slightly worse than much more complicated regularization techniques. It was also shown that the sampling methods of numerical inversion of the Laplace transform, discussed in Sec. VI C 1, are identical to the correlation method.[233]

The spectral function $g(\lambda)$ can be restored using the so-called ''frequency-scanned'' method. This method, as far as the authors know, was proposed independently by Henry *et al.*,[235] Ferenczi *et al.*[236] and Turchanikov *et al.*[237] In this method, the duration of the weighting function t_c is varied for each transient until a maximum in the dependence $S(\tau_s)$ is reached. An example of its application can be found in Ref. 238.

Correlation analysis can distinguish exponential components with a ratio of time constants $\tau_i/\tau_{i+1}=3.4$ if the SNR in the input transient is greater than 900.[234] It can also be applied to decays with SNR as low as 5, but in this case the resolution will not exceed $\tau_i/\tau_{i+1}\approx 15$.

3. Approximation by orthogonal functions

The method of approximation by a series of orthogonal functions is widely used in mathematics. If $g(\lambda)$ is represented by a series of orthogonal functions $\varphi_k(\lambda)$, then the Laplace integral equation, Eq. (5), will have the form

$$f(t)=L[g(\lambda)]=L\left(\sum_k a_k\varphi_k\right)=\sum_k a_k L(\varphi_k). \tag{75}$$

The coefficients a_k can be found using a simple numerical procedure, provided an analytical expression for $L(\varphi_k)$ is known. It has been proposed that we approximate $g(\lambda)$ by an infinite sum of orthogonal polynomials such as Legendre,[31,239–241] Laguerre,[30,239,242–246] Chebyshev[30,247–250] or Jacobi[251–254] polynomials, by trigonometric functions[239,241,255] or by Fourier series.[30] Davies *et al.*,[221] who tested 14 numerical methods of solving the Laplace integral equation, concluded that the use of Laguerre and Chebyshev polynomials gives very good accuracy over a wide range of functions. Furthermore, an approximation by orthogonal polynomials already includes regularization, since the requirement that the solution can be approximated by a polynomial of a given degree provides a filtering of solutions.

However, the most natural choices of an orthogonal set of functions for the inversion of the Laplace transform are singular functions (or eigenfunctions) and singular values (or eigenvalues) of the Laplace integral operator. Eigenfunctions and eigenvalues (see Smithies[55]) are determined for the case that the domains of $f(t)$ and $g(\lambda)$ coincide, and were discussed in Sec. IV above. If the domains of $f(t)$ and $g(\lambda)$ are different, a generalization of the eigenfunction expansion is provided by the singular function expansion.[256] The main formulas and the concept of the number of degrees of freedom and of the resolution ratio introduced in Sec. IV remain valid also for the singular value expansion. Various numerical techniques for evaluation of the singular values and singular functions are known, see, e.g., Refs. 58, 97. The eigenfunction expansion was practically employed by Provencher,[97] who used it to obtain an initial approximation for the least squares analysis in his program DISCRETE. However, as far as the authors know, the program DISCRETE is probably the only example of application of singular value decomposition in practical exponential analysis. Since the methods of Gardner transformation and Tikhonov regularization (discussed in the next sections) are much more commonly used, we will not go into further details of the method of approximation by orthogonal functions.

4. Fourier transform (Gardner transformation)

Gardner *et al.*[95,257] proposed the following solution of Eq. (5). The substitution $\lambda=\exp(-y)$ and $t=\exp(x)$ transforms the Laplace integral, Eq. (5), into a convolution integral:[258]

$$ff(x)=gg(x)\otimes kk(x)=\int_{-\infty}^{\infty} gg(y)kk(x-y)dy, \qquad (76)$$

where $gg(x)=g[\exp(-x)]$, $ff(x)=\exp(x)f[\exp(x)]$, and $kk(x)=\exp(x)\exp[\exp(x)]$. These types of integrals may be deconvolved using the Fourier transform technique:[259,260]

$$g(e^{-y})=\mathfrak{I}^{-1}\{\mathfrak{I}[e^x f(e^x)]/\mathfrak{I}[e^x\exp(-e^x)]\}, \qquad (77)$$

where $\mathfrak{I}$ is the Fourier transform operator. A graph of $g(e^{-y})$ as a function of $\lambda=e^{-y}$ will thus show maxima whenever $\lambda=\lambda_i$, with the amplitude proportional to A_i/λ_i.

The method of Gardner[95] did not attract much attention when proposed in 1959 because no effective algorithms for the computation of Fourier integrals were available. The technique was later modernized by Schlesinger[261] using the fast fourier transform (FFT) algorithm[262] to approximate the Fourier integral. An expression quite similar to Eq. (77) can also be obtained using Mellin transform (see Refs. 64, 263) instead of the Fourier transform; however, the computational methods for the FFT are much more effective than the algorithms used to calculate the Mellin transform.

The computation of Fourier transforms by numerical integration on an infinite interval is clearly impossible. Therefore, the integration interval must be truncated on both sides, i.e., the cutoff points $\pm x_0$ of the integral in the direct Fourier transform $\mathfrak{I}$ and the cutoff points $\pm\mu_0$ of the inverse Fourier transform $\mathfrak{I}^{-1}$ must be introduced. The cut-off procedure in $\mathfrak{I}$ usually results in spurious high-frequency components in the spectrum of the direct Fourier transform and in error ripples in the plot of $g(e^{-y})$ vs y, which tend to obscure the results. The best solution to decrease the amplitude of ripples would be to follow the transient for a sufficiently long time.[98] An appropriate choice of the cut-off frequency μ_0 of the inverse Fourier transform may also decrease the amplitudes of ripples. However, if μ_0 is chosen to be too small, there is an unnecessary loss of resolution of the peaks in the result of calculations. Thus, μ_0 serves as regularization parameter.

Provencher[98] described in detail a computer program for the realization of the Gardner transform. A number of practical details on how the best possible accuracy and resolution of Gardner's transform can be achieved, and examples of its practical applications, can be found in Refs. 98, 258, 264–273. Although the method is not sensitive to baseline offset if the integration in Eq. (77) is done on an infinite interval, in real-life applications where the integrals must be truncated (as discussed above), the stability and accuracy of the method will be higher if the baseline is removed. The resolution capacity of the Gardner transform was tested by many researchers. The number of points in the FFT varied from 32 to 256. All authors[73,95,257,258,261,264,265,266,271,273,274] came to the conclusion that the resolution limit for noise-free decays was between $\tau_1/\tau_2\sim 1.7$ and $\tau_1/\tau_2=2.5$. The signal-to-noise ratio SNR$\sim$200 decreases the resolution limit to $\tau_1/\tau_2\sim 4$ (Ref. 264), SNR$=$100 decreases the resolution to $\tau_1/\tau_2=5$ (Ref. 273). For noisy transients (SNR$=$20) (Ref. 273) even the number of components could not be reliably detected.

5. Tikhonov regularization method

The regularization method for solving the Fredholm integral equations of the first kind was proposed independently by Phillips[49] and in a more general form by Tikhonov,[275] see also monographs by Tikhonov[90] and Morozov.[276,277] In this method, instead of an exact solution, one searches for an approximate solution, which gives a minimum to the following functional:

$$M_\alpha[g(\lambda)]=\left\| \int_a^b K(t,\lambda)g(\lambda)d\lambda - f(t)\right\| + \alpha\Omega[g(\lambda)], \qquad (78)$$

where Ω is a regularizing (or smoothing) functional, and K is the exponential kernel. The practical implementation of Tikhonov's regularization requires one to choose a smoothing functional Ω suitable for a given problem and, for a given Ω, the regularization parameter α. Then, the solution $g(\lambda)$ is found using one of the minimization techniques. The following expressions for the regularizing functional are frequently used:[275]

$$\Omega[g(\lambda)]=\int_a^b\left(k(\lambda)\frac{dg(\lambda)}{d\lambda}+p(\lambda)g^2(\lambda)\right)d\lambda \tag{79}$$

and

$$\Omega[g(\lambda)]=\int_a^b\left[\sum_{i=0}^{n}K_i(\lambda)\left(\frac{d^{(i)}g(\lambda)}{d\lambda^{(i)}}\right)^2\right]d\lambda, \tag{80}$$

where $k(\lambda)$ and $p(\lambda)$ are smooth positive functions. Frequently, only the second derivative of $g(\lambda)$ is used in Eq. (80):

$$\Omega[g(\lambda)]=\int_a^b\left(\frac{d^2g(\lambda)}{d\lambda^2}\right)^2 d\lambda. \tag{81}$$

The regularization parameter α determines the balance between the exact and a smoothed solution of Eq. (5). The solution which gives a minimum to Eq. (78) for $\alpha=0$ may have no physical sense and may be unstable with respect to small changes in $f(t)$. On the other hand, if the parameter α is large, the solution will be both smooth and stable, but a significant part of the physical information contained in $f(t)$ will be lost.[278] Several methods were proposed to choose the optimum α. Most of them depend on *a prior* knowledge of the noise level in the data.[90,278–283] It should be noted that in practical cases, when the noise level is known only approximately, the selection of α by means of an automated computer algorithm remains a challenging task. An interesting method, which gives a clear understanding of how the regularization works, chooses α using the so-called L curve.[106,108,284–286] A survey of methods for selection of the parameter α was made by Davies.[287] The problems of existence of a solution when Tikhonov regularization is used, its uniqueness, convergence rate, etc., were widely discussed in the mathematical literature, and the reader may refer to Refs. 90, 277, 288–294.

Equation (78) may be solved by any optimization technique, for example by the NLS technique. However, it was shown to be computationally effective[90,289] to search $g(\lambda)$ as the solution of the Euler equation for the functional M_α, determined in Eq. (78):

$$\nabla M_\alpha[g(\lambda)]=0. \tag{82}$$

Several computer programs that utilize Tikhonov regularization algorithms are known. Provencher[48] distributed his program CONTIN[295] to more than a hundred research laboratories. His program was employed for high-resolution capacitance spectroscopy of semiconductors by Morimoto *et al.*,[296,297] Fudamoto *et al.*,[298] Tahira *et al.*,[299] Yoshino *et al.*,[300] Maier *et al.*,[301] Batovski *et al.*,[302,303] and by Dobaczewski *et al.*[72,304,305] The Tikhonov regularization method was also implemented by Weese (program FTIKREG[306,307] and program NLREG[308]). Riele suggested the program F1REGU.[309] One should note that despite the apparent simplicity of the Tikhonov regularization, it requires rather complicated computational algorithms to accommodate numerical integration, differentiation, selection of the regularization parameter and minimization technique. The programs CONTIN and FTIKREG, mentioned above, contain about 5000 lines of the FORTRAN code. It is a very time-consuming task to write such a program, and we strongly recommend using one of the available programs rather than to write it oneself. To the best of our knowledge, the code of the above mentioned programs is available from the CPC program library.[310] The program CONTIN can also be downloaded from the Internet.[311]

Tsema[312] tested the Tikhonov regularization algorithm for a double exponential and showed that two exponentials with $\tau_1/\tau_2>2$ can be reliably resolved if SNR$\geqslant 10^3$. Tarasov[313] showed that two exponentials with $\tau_2/\tau_1=3$ can be distinguished with SNR$=100$, while exponentials with $\tau_2/\tau_1=5$ can be distinguished for SNR>15.

6. Method of maximum entropy

The method of maximum entropy stems from probability theory and information theory, which introduce a criterion for the amount of uncertainty represented by a discrete probability distribution $(p_1\cdots p_n)$. Clearly, a measure of randomness, or prior uncertainty H of the data, should have a zero value when the number of probabilities is unity, i.e., when there is no spread at all, and should be positive when there is more than one probability. Furthermore, it should be additive for independent trials, i.e., $H(pq)=H(p)+H(q)$. Obviously, the logarithm $\log(n)$, among many other functions, satisfies these requirements.[53,54] Since it is just the expression for entropy as found in statistical mechanics, it was called the entropy of the probability distribution, or simply the entropy. Henceforth one can consider the terms ''entropy'' and ''uncertainty'' as synonymous.

The entropy, as a measure of uncertainty of the outcome of an experiment, is largest when all admissible outcomes have equal probabilities. The principle of maximum entropy is very simple: when making inferences on the basis of a partial information we must use that probability distribution which has maximum entropy subject to whatever is known. This is the only unbiased assignment we can make. This maximum entropy description retains all of the uncertainty not removed by the data, and thus it tends to be most objective or maximally noncommittal with respect to missing information.[314] More details on the basic principles of maximum entropy can be found in Jaynes,[315–321] Skilling,[322] Gray[323] and Aczel *et al.*[324] A very useful introduction into information theory and the concept of entropy is presented by Woodward.[325]

In experimental science one deals with the values of certain physical parameters rather than with their probabilities, and the method of maximum entropy (MEM) is used primarily as a regularization technique with a regularizing functional identical or similar to the entropy term.[326] It is supposed that the unknown function $g(\lambda)$ has the properties of the probability distribution [i.e., $g(\lambda)>0$ and, when normal-

ized, $\int g(\lambda)d\lambda = 1$], and the data $f_k = f(t_k)$ are considered as constraints. The objective is to choose from all feasible solutions $g(\lambda)$ the one that maximizes the entropy functional, $-\int g(\lambda)\log(\lambda)d\lambda$. This problem can be solved algebraically using the method of Lagrange multipliers,[326,327] or as a combination of the principle of maximum entropy and the algorithm of Tikhonov regularization:[328–331]

$$M[g(\lambda)] = \|Kg - f\| + \alpha \left\| g(\lambda)\log\left(\frac{g(\lambda)}{m(\lambda)}\right)\right\|, \tag{83}$$

where α can be considered as a Lagrange multiplier or as a regularization parameter, and $m(\lambda)$ is a prior estimate of $g(\lambda)$. Landl *et al.*[332] discussed regularization functionals of the form:

$$\Omega(g) = \int \mathbf{D}(g)\ln(\mathbf{D}g)d\lambda, \tag{84}$$

where **D** denotes an ordinary differential operator with constant coefficients of the form:

$$\mathbf{D}g = \sum_{p=0}^{P} a_p \frac{d^p g}{d\lambda^p}. \tag{85}$$

MEM-based filtering is used in physics primarily for image processing[333] such as forensic imaging, radio astronomy, medical tomography, and plasma tomography. The image restoration techniques benefit from the property of the MEM functionals $\ln(g)$ and $-g\ln(g)$ to give reconstructions that are positive, have sharpened peaks and flattened baselines.[334] An application of the MEM principle to the solution of the Laplace integral equation was discussed by Gzyl,[335] which is so far the only application of MEM to exponential analysis known to the authors.

VII. COMPARISON OF EXPONENTIAL ANALYSES

The idea of comparing numerical methods for exponential analysis and choosing the one with the best resolution has always been very attractive. A standard procedure to compare different algorithms is to apply them to one or several simulated sets of input data (in our case, exponential decays) to compare the accuracy of the solution, the sensitivity to noise in the input data, and the time required to compute the solutions. Many scientists, e.g., McKinnon *et al.*,[336] Smith *et al.*,[73] O'Connor *et al.*,[337] Tittelbach-Helmrich,[193] Kirchner *et al.*,[82] Isenberg,[171] Bromage,[172] Thomasson *et al.*,[71] Zhang *et al.*,[75] Apanasovich *et al.*,[182] and Doolittle *et al.*[338,339] tested different fitting methods applying them to simulated or experimentally obtained relaxations. Their results revealed that the comparison procedure which works fine for well-posed problems, returns by far less conclusive results when one has to deal with the ill-posed problem of exponential analysis. Well-posed problems always have a solution, and the solution is unique. If the problem of exponential analysis was well posed, then the analysis of any transient would return the set of parameters, which was used to simulate the input data (assuming that the algorithm is correct). This is not the case with ill-posed problems. The solution is not unique, and the algorithm has to use certain assumptions to select one of the possible solutions.

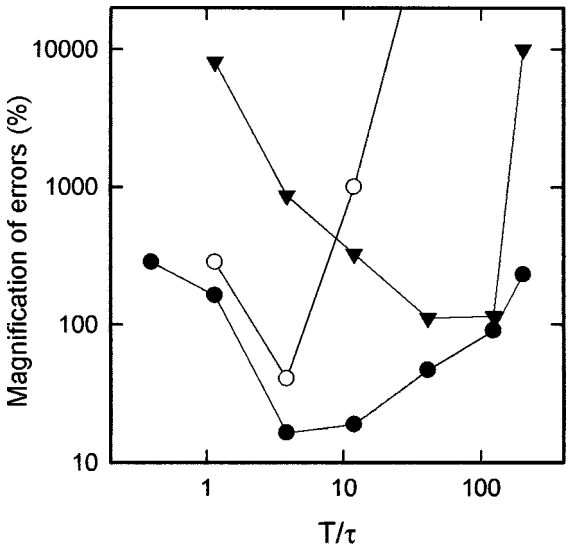

FIG. 8. Magnification of errors due to random noise in the input transients for three different algorithms (after Bromage, Ref. 172): nonlinear least squares analysis (filled circles) (Sec. VI B 2), algebraic method as was suggested by Prony (Ref. 29) (triangles) and in modification of Cornell (Ref. 340) (open circles) (Sec. VI A 2). The data are plotted as a function of the ratio of data acquisition time T to the mean decay time constant τ. Calculations were done for a double-exponential decay.

These assumptions vary from method to method, and selected by exponential analysis the ''best'' solution depends on a number of factors such as: the noise level in transients (and the distribution of errors in the simulated noise), the presence of the offset (and the accuracy of its determination and subtraction), the number of exponential components as well as the ratio of their time constants and amplitudes, and the ratio of the time constant of the slowest component to the measurement time.

For example, Bromage[172] analyzed three methods: two modifications of Prony's algebraic method (the method as it was proposed by Prony[29] and in modification of Cornell[340]) and the NLS method. He simulated double-exponential decays with equal amplitudes and the ratio of time constants of the components $\tau_2/\tau_1 = 2$. The decays with added random noise were analyzed using each of the three algorithms, and the scatter of retrieved parameters was compared with the known input noise level. The dependencies, represented in Fig. 8, were found to be parabola-like curves with a sharp minimum, which had an individual position for each method. The error in the determination of τ could be an order of magnitude higher than at the point of minimum, if T/τ was changed by a factor of 2 from the optimum value (see Fig. 8).

The strong dependence of the errors in Prony's method on parameters of the transients was later confirmed by detailed studies of Sun *et al.*[86,185] Consequently, a comparison of the methods for exponential analysis will be comprehensive only if: (a) all methods are tested using the same input data, and (b) the test includes transients with varied number of decay components, ratios of amplitude and decay rates of the components, noise level, baseline offset, ratio of decay acquisition time to the decay time constant T/τ, and if applicable, initial approximation for the fit. Such a comparison

would require a very large number of fits and is beyond the scope of this review article. Clearly, the literature analyses also do not comply with the requirements listed above. First, usually only two or three methods were tested by each author. Furthermore, different authors used entirely different sets of transients to test the methods, which often led to contradictory results (as it can be expected from the discussion of Fig. 8). Just to mention a few examples of such contradictions, Isenberg[171] demonstrated that if a decay consists of several components with similar time constants, even very small errors will cause the NLS fit to fail, whereas Grinvald *et al.*[47] came to the opposite conclusion. Tittelbach-Helmrich[193] concluded, that the method of moments gave large systematic errors even if the SNR of the decay was about 200 and excluded it from the analysis, while Kirchner *et al.*[82] reported that the method of moments worked fine even for SNR=10.

Despite disagreement in the literature about comparative performance of different methods, a very important conclusion about the resolution limit of all methods discussed above can be made. As discussed in Sec. IV, there is a fundamental resolution limit to exponential analysis, determined by the SNR in the input transients. According to Table I, the maximum resolution which can be achieved for the infinite domain of the solution is 2.44 for SNR=100, 1.88 for SNR=1000, and 1.63 for SNR=10^4. The SNR in experimental transients depends on the type of experiment and on the number of averaged transients, and seldom exceeds SNR=10^3–10^4. Therefore, the maximum resolution that can be reached in experimental setups, τ_{i+1}/τ_i, is about 1.6–1.8. The analysis of literature data reveals that this resolution has been reached using nearly all fitting and spectroscopic methods discussed in this review article. For NLS analysis (Sec. VI B 2), the resolution of $\tau_1/\tau_2>2$ was reported for SNR$\geqslant$1000 by Grinvald,[167] Morimoto *et al.*,[168,169] and Tahira *et al.*[170] The resolution of $\tau_1/\tau_2=1.5$ for a double exponential and $\tau_1/\tau_2=2$ for a triple exponential for simulated decays without noise was reported by Clayden.[214] Using the integration method (Sec. VI B 5), Tittelbach-Helmrich[193] resolved components in a double-exponential decay with $\tau_1/\tau_2=2.5$ even for the SNR as low as 30. The method of modulating functions (Sec. VI B 4) could resolve two exponentials with a ratio of time constants up to 2 (Refs. 191, 192). The same resolution was reported for the Prony's method[86] (Sec. VI B 3) and for the method of moments[171] (Sec. VI B 6). The Gaver–Stehfest sampling method (Sec. VI C 1) was tested by Nolte *et al.*,[222] who showed that without noise it can resolve two exponentials with $\tau_1/\tau_2=1.5$. The Gardner transform (Sec. VI C 4) was tested by a number of researchers,[73,95,257,258,261,264–266,271,273,274] who agreed that the resolution limit of the Gardner transform for noise-free transients is between $\tau_1/\tau_2\sim1.7$ and $\tau_1/\tau_2=2.5$. Tsema[312] and Tarasov[313] tested the Tikhonov regularization method (Sec. VI C 5) and reported a resolution of $\tau_1/\tau_2=2$ for SNR=1000 and $\tau_2/\tau_1=3$ for SNR=100. The resolution of correlation analysis (Sec. VI C 2) was reported to be $\tau_i/\tau_{i+1}=3.4$ for the SNR of about 900 (Ref. 234).

Thus, the problem of reaching the fundamental resolution limit is, in principle, solved for nearly all of the discussed methods (for some of the methods no test results are available). Hence, the resolution of the method is not a criterion to prefer one of the methods to another. Indeed, it is very important to choose the method suitable for a given experimental problem, such as for monoexponential or multiexponential analysis, or a spectroscopic method. This classification of methods was discussed in detail throughout the review. However, even if the required type of the algorithm is clear, there are a number of algorithms to choose from in each of these three groups. To make a final choice, the following factors must be taken into account: (a) the stability of each method with respect to a wide range variation of parameter values (in other words, how much the sensitivity of the method to noise and its resolution varies as the parameters of exponential decay vary); (b) the time it takes to compute a solution; (c) the difficulty of programming the algorithm and whether the program code is available from program libraries; (d) the tolerance of the method to a baseline offset.

The stability of the algorithms as the parameters are varied is an increasingly important problem. As discussed above, a comprehensive comparison of exponential analyses requires a large number of fits and was beyond the scope of this review article. The literature data are also insufficiently conclusive. In this review article we wish to emphasize this problem and hope that this discussion will stimulate detailed studies in the future.

The time required to compute a solution is gradually losing its importance as computers are getting faster. Yet the calculation time may in some cases remain an important factor to consider, since it may vary by several orders of magnitude. Generally, the computation time is largest when the variational methods are involved. It is worth noting that all methods (and particularly the methods discussed in Sec. VI B) can be divided into two groups: methods based on variation methods, like nonlinear least squares; and methods which are based on systems of algebraic equations. The first group of methods anticipates that the transient can be described by Eq. (4) and varies parameters until a good agreement between the experimental transient and Eq. (4) is achieved. The other group of methods (Prony's method, differentiation method, integration method, method of moments, Laplace–Padé approximation) is based on a certain equation which includes the decay $f(t)$ (and eventually its derivatives, moments, etc.) and time constants and amplitudes of the decay. This equation is then transformed into a system of equation by substituting $f(t)$ with experimental data points $f_{\exp}(t_i)$, $i=1,2,\ldots,N$, and the unknown parameters are determined by solving this system of equations. These methods are usually much faster, although some of them are also more sensitive to noise.

Programming of methods of monoexponential and multiexponential analysis is quite straightforward. The algorithms for least squares minimization, FFT, and the solution of matrix equations, which are parts of some exponential analyses, are well documented and the code of these subroutines can be found in textbooks (see, e.g., Ref. 100). Some of the spectroscopic methods, such as the sampling method or correlation method, are also very easy to program. On the

other hand, the mathematics involved in most powerful spectroscopic methods such as Tikhonov regularization and Gardner transform is quite complicated. Fortunately, the program codes of these programs are readily available from the CPC program library.[310] Some of the programs can be downloaded from the Internet.[311]

Tolerance of the method to baseline offsets may be very important, depending on whether a baseline offset can be encountered in the experiment. Few of the discussed techniques are tolerable of nonzero baseline offsets. The correlation method (Sec. VI C 2) is probably the only algorithm which is absolutely baseline insensitive. The monoexponential Fourier transform of the transients (Sec. VI A 1), Gardner transformation (Sec. VI C 4), the NLS method (Sec. VI B 2), integration method (Sec. VI B 5), Tikhonov regularization (Sec. VI C 5) and the method of maximum entropy (Sec. VI C 6) can accommodate the baseline offset as an additional parameter, but this usually results in a higher sensitivity to noise in the input transients than in the case without offset. Some of the sampling methods (Sec. VI C 1) are also tolerant to baseline offsets. However, such commonly used methods as Prony's method (Sec. VI B 3), or method of moments (Sec. VI B 5) will provide wrong results or will even crash if a decay contains an offset. Unfortunately, algorithms for extrapolation of baseline offsets are poorly developed and are not always sufficiently exact. Therefore, we would recommend algorithms that do not require baseline corrections and can accommodate transients with a baseline.

If we put together all pieces of information on each particular method as they are available from the literature, taking into account the features which we expect each method to have, including stability for wide range changes in parameters and insensitivity to baseline offsets, then the best methods (according to the opinion of these authors) would be: Fourier transform of the decays (Sec. VI A 1) for the methods for monoexponential analysis; the NLS method (Sec. VI B 2) for the fitting methods, and correlation method (Sec. VI C 2) and Tikhonov regularization (Sec. VI C 5) for the spectroscopic methods. It should be noted that the correlation method is extremely simple to program and is very versatile, although the Tikhonov regularization enables one to get the most of resolution for a given SNR in the transient.

VIII. DISCUSSION

We have reviewed an extensive number of articles from various branches of experimental physics and mathematics, discussed general limitations of exponential analysis and, finally, summarized the algorithms for exponential analysis suggested in the literature. At first glance, it appears that there is a great variety of numerical algorithms for exponential analysis. However, a deeper insight in their principles and classification of the algorithms, suggested in this review article, shows that all methods use the same fundamental regularization principle, based either on a model of the function $g(\lambda)$ (as in fitting methods, Secs. VI A and VI B) or on an assumption of its smoothness (as in spectroscopic methods, Sec. VI C). In some cases, a direct correlation between the regularization parameters of two different spectroscopic methods has been derived analytically (see, e.g., Refs. 341, 342). We believe that the understanding of this fundamental principle and of the principle limitations of exponential analysis is of major importance for every scientist involved in evaluations of exponential decays.

The main parameter of exponential analysis is its resolution, i.e., the ratio of the time constants of two exponentials which can be resolved in the transient. It is very important to bear in mind (see Sec. IV) that there is a principle limitation of resolution of exponential analysis. This limitation is inherent in the nature of the problem itself and cannot be improved by developing new algorithms. The resolution limit is determined by the SNR in transients (see Table I and Fig. 4 in Sec. IV). According to the literature, most of the methods discussed in this review article have reached the resolution close to the resolution limit of exponential analysis for a given SNR. There are indications that the major trend in development of programs for exponential analysis is now shifting from achievement of the highest resolution possible to development of programs that are independent of initial approximations, stable with respect to variations of time constants and amplitudes of exponential components, and insensitive to baseline offsets. Frequently, a combination of several techniques is required to satisfy these requirements. The fitting routines impose stronger regularization on the problem and work well even for high signal-to-noise ratios. However, they are accurate only if the hypothesis of the number of components is correct and the initial approximation is close to the true solution. A way to obtain this initial approximation is to extract it from a spectroscopic method, as it was done by Provencher[97] and Mazzola *et al.*[343] Daniels[129] pointed out that many programs based on the NLS method frequently contain more than one technique. These programs first use a ''slow but sure'' algorithm such as the simplex to find a good initial approximation, then a faster technique when sufficiently close to the minimum.

Although the problem of acquisition of transients is often neglected in favor of computer algorithms for their analysis, the SNR of the transients and, consequently, the resolution of the exponential analysis are to a great extent determined by the experimental setup. As it was discussed in Sec. V, an incorrect choice of the settings of the analog-to-digital converter, too short duration of the digitized transient, or instabilities of the experimental equipment during the data accumulation can impose severe limitations on the resolution capacity of exponential analysis.

Finally, it is worth noting that complicated numerical techniques are not always the only way to resolve several exponents with close time constants. In fact, closely adjacent exponential components can frequently be resolved by changing the experimental conditions of the measurement, for example the temperature, or intensity of excitation, etc. A number of examples of this approach can be found in the literature.[338,344]

ACKNOWLEDGMENTS

The authors are grateful to S. Foner for discussions which helped to determine the subject of this review article,

L. Dobaczewski for sharing his experience on high-resolution DLTS, C. Flink, H. Hieslmair, K. Knobloch, and W. Seifert for a critical reading of the manuscript, and E. R. Weber for overall support of this work.

[1] A. K. Zakrzewski, L. Dobaczewski, G. Karczewski, T. Wojtowicz, and J. Kossut, Acta Phys. Pol. A **88**, 961 (1995).
[2] B. B. Mittal, V. Sathiaseelan, A. W. Rademaker, and M. C. Pierce, Int. J. Radiat. Oncol., Biol., Phys. **21**, 1353 (1991).
[3] V. Guardabasso, P. J. Munson, and D. Rodbard, Comput. Methods Programs Biomed. **27**, 55 (1988).
[4] S. F. J. Langer, Comput. Methods Programs Biomed. **53**, 57 (1997).
[5] M. J. Ruiz, Am. J. Phys. **55**, 641 (1987).
[6] N. Ohtomo, T. Kamo, M. Watanabe, K. Yoneyama, Y. Tanaka, and R. Hayashi, Jpn. J. Appl. Phys., Part 1 **35**, 5571 (1996).
[7] A. F. Oberhauser, P. E. Marszalek, H. P. Erickson, and J. M. Fernandez, Nature (London) **393**, 181 (1998).
[8] G. Chirico and G. Baldini, J. Mol. Liq. **41**, 327 (1989).
[9] D. J. Sorce, R. R. Sciacca, and A. M. Keller, J. Mater. Res. **87**, 230 (1990).
[10] D. Liu, P. A. Knauf, and S. D. Kennedy, Biophys. J. **70**, 715 (1996).
[11] K. J. Willis, W. Neugebauer, M. Sikorska, and A. G. Szabo, Biophys. J. **66**, 1623 (1994).
[12] Z. Bajzer, and F. G. Prendergast, Biophys. J. **65**, 2313 (1993).
[13] J. Wendler and A. R. Holzwarth, Biophys. J. **52**, 717 (1987).
[14] T. M. Yoshida and B. G. Barisas, Biophys. J. **50**, 41 (1986).
[15] G. W. Rayfield, Biophys. J. **48**, 111 (1985).
[16] J. R. Lindroth, S. M. Martin, and J. W. Ledbetter, Comput. Biol. Med. **17**, 369 (1987).
[17] D. Sonnenschmidt and K. H. Vanselow, Meas. Sci. Technol. **7**, 1536 (1996).
[18] W. E. Schulz, H. D. Smith, Jr., J. L. Verbout, J. R. Bridges, and G. H. Garcia, Nucl. Instrum. Methods Phys. Res. B **10–11**, 1023 (1985).
[19] F. Rocadenbosch, A. Comeron, and D. Pineda, Appl. Opt. **37**, 2199 (1998).
[20] M. Muraoka and H. Abé, J. Am. Ceram. Soc. **79**, 51 (1996).
[21] R. Richert, Chem. Phys. Lett. **118**, 534 (1985).
[22] P. Marshall, Comput. Chem. **11**, 219 (1987).
[23] J. P. Brody and P. Yager, Sens. Actuators A **58**, 13 (1997).
[24] M. A. Daeubler, G. W. Warren, I. M. Bernstein, and A. W. Thompson, Metall. Trans. A **22**, 521 (1991).
[25] D. Goncalves, B. Matvienko, and L. O. S. Bulhoes, Synth. Met. **83**, 147 (1996).
[26] G. Rocchini, Corros. Sci. **30**, 9 (1990); **33**, 1759 (1992).
[27] M. Hajduga and J. Kucera, Acta Technica CSAV **42**, 741 (1997).
[28] A. Martin, A. G. Casielles, M. G. Munoz, F. Ortega, and R. G. Rubio, Phys. Rev. E **58**, 2151 (1998).
[29] R. de Prony, J. Ec. Polytech. **1**, 24 (1795).
[30] C. Lanczos, *Applied Analysis* (Prentice-Hall, Englewood Cliffs, NJ, 1959), p. 272 ff.
[31] R. E. Bellman, R. E. Kalba, and J. Lockett, *Numerical Inversion of the Laplace Transform* (Elsevier, New York, 1966).
[32] D. V. Lang, *Thermally Stimulated Relaxation in Solids*, edited by P. Braunlich in Topics In Applied Physics, Vol. 37 (Springer, Berlin, 1979).
[33] G. L. Miller, D. V. Lang, and L. C. Kimerling, Annu. Rev. Mater. Sci. **7**, 377 (1977).
[34] D. V. Lang, J. Appl. Phys. **45**, 3014 and 3023 (1974).
[35] A. A. Istratov and O. F. Vyvenko, Semiconductors **29**, 340 (1995).
[36] G. M. Martin, A. Mitonneu, and A. Mircea, Electron. Lett. **13**, 191 (1977).
[37] K. Yamanaka, S. Naritsuka, K. Kanamoto, M. Mihara, and M. Ishii, J. Appl. Phys. **61**, 5062 (1987).
[38] G. Doetsch, *Handbuch der Laplace-transformation* (Birkhäuser, Basel, 1950) (in German).
[39] F. G. Tricomi, *Integral Equations* (Interscience, New York, 1957).
[40] G. Wahba, in: *Solution Methods for Integral Equations: Theory and Applications*, edited by M. A. Goldberg (Plenum, New York, 1978).
[41] M. Bertero, P. Boccacci, and F. Maggio, Int. J. Imaging Syst. Technol. **6**, 376 (1995).
[42] A. Pinkus and S. Zafrany, *Fourier Series and Integral Transforms* (Cambridge University Press, Cambridge, 1997), p. 98.
[43] F. Oberhettinger, *Tabellen zur Fourier Transformation* (Springer, Berlin, 1957), pp. 10, 121.
[44] E. Hille, *Analysis* (Blaisdell, Waltham, MA, 1966), Vol. 2, p. 126.
[45] A. C. Pipkin, *A Course of Integral Equations* (Springer, New York, 1991), p. 72.
[46] R. S. Julius, Comput. Biomed. Res. **5**, 473 (1972).
[47] A. Grinvald and I. Z. Steiberg, Anal. Biochem. **59**, 583 (1974).
[48] S. W. Provencher, Comput. Phys. Commun. **27**, 213 (1982).
[49] D. L. Phillips, J. ACM **9**, 84 (1962).
[50] V. K. Ivanov, USSR Comput. Math. Math. Phys. **6**, 197 (1966).
[51] H. Nyquist, Bell Syst. Tech. J. **3**, 324 (1924).
[52] H. Nyquist, A.I.E.E. Trans. **47**, 617 (1928).
[53] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, IL, 1949).
[54] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948); **27**, 623 (1948).
[55] F. Smithies, *Integral Equations* (Cambridge University Press, Cambridge, 1958).
[56] J. C. McWhirter and E. R. Pike, J. Phys. A **11**, 1729 (1978).
[57] V. V. Petrov and A. S. Uskov, Sov. Math. Dokl. **11**, 1571 (1970).
[58] E. R. Pike, J. G. McWhirter, M. Bertero, and C. De Mol, Proc. IEEE **131**, 660 (1984).
[59] M. Bertero, P. Boccacci, and E. R. Pike, Proc. R. Soc. London, Ser. A **383**, 15 (1982).
[60] M. Bertero, P. Boccacci, and E. R. Pike, Proc. R. Soc. London, Ser. A **393**, 51 (1984).
[61] M. Bertero, P. Brianzi, and E. R. Pike, Proc. R. Soc. London, Ser. A **398**, 23 (1985).
[62] M. Bertero and C. De Mol, Atti Fond. Giorgio Ronchi **36**, 619 (1981).
[63] M. Bertero, C. De Mol, and E. R. Pike, Inverse Probl. **4**, 573 (1988).
[64] M. Bertero and E. R. Pike, Inverse Probl. **7**, 1 (1991); **7**, 21 (1991).
[65] M. Bertero, F. A. Grünbaum, and L. Rebolia, Inverse Probl. **2**, 131 (1986).
[66] N. Ostrowsky, D. Sornette, P. Parker, and E. R. Pike, Opt. Acta **28**, 1059 (1981).
[67] A. J. Jerri, Proc. IEEE **65**, 1565 (1977).
[68] M. Bertero and E. R. Pike, Opt. Acta **29**, 727 (1982).
[69] M. Bertero, P. Brianzi, and E. R. Pike, Inverse Probl. **1**, 1 (1985).
[70] F. R. Shapiro, S. D. Senturia, and D. Adler, J. Appl. Phys. **55**, 3453 (1984).
[71] W. M. Thomasson and J. W. Clark, Math. Biosci. **22**, 179 (1974).
[72] L. Dobaczewski, P. Kaczor, I. D. Hawkins, and A. R. Peaker, J. Appl. Phys. **76**, 194 (1994).
[73] M. R. Smith, S. Cohn-Sfetcu, and H. A. Buckmaster, Technometrics **18**, 467 (1976).
[74] P. Hall and B. Selinger, J. Phys. Chem. **85**, 2941 (1981).
[75] Z. Zhang, K. T. V. Grattan, Y. Hu, A. W. Palmer, and B. T. Meggitt, Rev. Sci. Instrum. **67**, 2590 (1996).
[76] L. Dobaczewski (private communications).
[77] W. R. Bennett, Bell Syst. Tech. J. **27**, 446 (1948).
[78] M. P. Ekstrom, IEEE Trans. Audio Electroacoust. **AU-21**, 344 (1973).
[79] W. A. Doolittle and A. Rohatgi, Rev. Sci. Instrum. **63**, 5733 (1992).
[80] P. V. Kolev, M. J. Deen, and N. Alberding, Rev. Sci. Instrum. **69**, 2464 (1998).
[81] L. J. Dowell and G. T. Gillies, Rev. Sci. Instrum. **62**, 242 (1991).
[82] P. D. Kirchner, W. J. Schaff, G. N. Maracas, L. F. Eastman, T. I. Chappell, and C. M. Ransom, J. Appl. Phys. **52**, 6462 (1981).
[83] K. Ikossi-Anastasiou and K. P. Roenker, J. Appl. Phys. **61**, 182 (1987).
[84] W. S. Moore and T. Yalcin, J. Magn. Reson. **11**, 50 (1973).
[85] M. R. Smith and H. A. Buckmaster, J. Magn. Reson. **17**, 29 (1975).
[86] T. Sun, Z. Y. Zhang, K. T. V. Grattan, and A. W. Palmer, Rev. Sci. Instrum. **68**, 58 (1997).
[87] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965), pp. 883, 914.
[88] P. C. Mangelsdorf, J. Appl. Phys. **30**, 442 (1959).
[89] W. Perl, Int. J. Appl. Radiat. Isot. **8**, 211 (1960).
[90] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems* (Winston/Wiley, New York, 1977).
[91] T. F. Dolgopolova and V. K. Ivanov, USSR Comput. Math. Math. Phys. **6**, 570 (1966) (in Russian).
[92] M. Bertero, C. De Mol, and E. R. Pike, Inverse Probl. **1**, 301 (1985).
[93] R. Migneron, K. S. S. Narayanan, and H. Rasmussen, Comput. Phys. Commun. **54**, 239 (1989).
[94] I. Isenberg, R. D. Dyson, and R. Hanson, Biophys. J. **13**, 1090 (1973).
[95] D. G. Gardner, J. C. Gardner, G. Laush, and W. W. Meinke, J. Chem. Phys. **31**, 978 (1959).
[96] R. D. Dyson and I. Isenberg, Biochem. J. **10**, 3233 (1971).
[97] S. W. Provencher, J. Chem. Phys. **64**, 2772 (1976).
[98] S. W. Provencher, Biophys. J. **16**, 27 (1976).

[99] C. M. Rader and B. Gold, in *Digital Signal Processing*, edited by L. R. Rabiner and C. M. Rader (IEEE, New York, 1972).
[100] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge, 1992).
[101] M. R. Smith and S. Cohn-Sfetcu, J. Phys. E **8**, 515 (1975).
[102] A. C. Kokaram, N. Persad, J. Lasenby, W. J. Fitzgerald, A. McKinnon, and M. Welland, Appl. Opt. **34**, 5121 (1995).
[103] G. R. Arce, IEEE Trans. Signal Process. **39**, 1146 (1991).
[104] A. Nieminen, P. Heinonen, and Y. Nuevo, IEEE Trans. Pattern. Anal. Mach. Intell. **PAMI-9**, 74 (1987).
[105] F. Paisan and A. Crise, Signal Process. **6**, 67 (1984).
[106] P. C. Hansen, Inverse Probl. **8**, 849 (1992).
[107] P. C. Hansen, Numer. Algorithms **6**, 1 (1994).
[108] P. C. Hansen, SIAM (Soc. Ind. Appl. Math.) J. Sci. Stat. Comput. **11**, 503 (1990).
[109] J. M. Varah, SIAM (Soc. Ind. Appl. Math.) J. Sci. Stat. Comput. **4**, 164 (1983).
[110] K. Ikeda and H. Takaoka, Jpn. J. Appl. Phys., Part 1 **21**, 462 (1982).
[111] C. D. Wang and H. C. Lin, J. Appl. Phys. **52**, 546 (1981).
[112] S. Weiss and R. Kassing, Solid-State Electron. **31**, 1733 (1988).
[113] M. Okuyama, H. Takakura, and Y. Hamakawa, Solid-State Electron. **26**, 689 (1983).
[114] P. D. Devries and A. A. Khan, J. Electron. Mater. **18**, 543 (1989).
[115] P. D. Devries and A. A. Khan, J. Electron. Mater. **18**, 763 (1989).
[116] F. Hasegawa, Jpn. J. Appl. Phys., Part 1 **24**, 1356 (1985).
[117] C. A. B. Ball and A. B. Conibear, Mater. Sci. Forum **83–87**, 1147 (1992).
[118] T. Mukoyama, Nucl. Instrum. Methods Phys. Res. **197**, 397 (1982).
[119] V. I. Zubkov, A. V. Solomonov, and M. T. Todorov, Sov. Phys. Semicond. **21**, 1050 (1987).
[120] H. K. Kim, T. E. Schlesinger, and A. G. Milnes, J. Electron. Mater. **17**, 187 (1988).
[121] H. D. Van Liew, Science **138**, 682 (1962).
[122] H. D. Van Liew, J. Theor. Biol. **16**, 43 (1967).
[123] P. Mancini and A. Pilo, Comput. Biomed. Res. **3**, 1 (1970).
[124] P. R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences* (McGraw-Hill, New York, 1969).
[125] H. H. Rosenbrock, Comput. J. (UK) **3**, 175 (1960).
[126] M. J. D. Powell, Comput. J. (UK) **7**, 155 (1964).
[127] R. Fletcher, Comput. J. (UK) **8**, 33 (1965).
[128] J. A. Nelder and R. Mead, Comput. J. (UK) **7**, 308 (1965).
[129] R. W. Daniels, *An Introduction to Numerical Methods and Optimization Techniques* (North-Holland, New York, 1978), p. 177 ff.
[130] W. Spendley, G. R. Hext, and F. R. Himsworth, Technometrics **4**, 441 (1962).
[131] R. Fletcher and C. M. Reeves, Comput. J. (UK) **7**, 149 (1964).
[132] J. W. Daniel, *The Approximate Minimization of Functionals* (Prentice-Hall, Englewood Cliffs, N.J., 1971).
[133] G. H. Golub and D. P. O'Leary, SIAM (Soc. Ind. Appl. Math.) Rev. **31**, 50 (1989).
[134] C. G. Broyden, Math. Comput. **21**, 368 (1967).
[135] J. G. P. Barnes, Comput. J. (UK) **8**, 66 (1965).
[136] C. G. Broyden, Math. Comput. **19**, 577 (1965).
[137] R. Fletcher and M. J. D. Powell, Comput. J. (UK) **6**, 163 (1963).
[138] P. Wolfe, Commun. ACM **2**, 12 (1959).
[139] W. C. Davidon, Comput. J. (UK) **10**, 406 (1968).
[140] S. L. Laiken and M. P. Printz, Biochemistry **9**, 1547 (1970).
[141] K. C. Woo and L. C. Chow, Numer. Heat Transfer **4**, 499 (1981).
[142] G. H. Golub and V. Pereyra, SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **10**, 413 (1973).
[143] J. Pitha and R. C. Jones, Can. J. Chem. **6**, 3031 (1966).
[144] D. W. Marquardt, SIAM (Soc. Ind. Appl. Math) J. **11**, 431 (1963).
[145] J. E. Dennis and R. E. Welsch, Commun. Statist. **B7**, 345 (1978).
[146] Y. Bard, SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **7**, 157 (1970).
[147] M. J. Box, Comput. J. (UK) **9**, 67 (1966).
[148] J. B. Crockett and H. Chernoff, Pacific J. Math. **5**, 33 (1955).
[149] J. Greenstadt, Math. Comput. **21**, 360 (1967).
[150] H. A. Spang, SIAM (Soc. Ind. Appl. Math.) Rev. **4**, 343 (1962).
[151] J. Meiron, J. Opt. Soc. Am. A **55**, 1105 (1965).
[152] P. D. Flanagan, P. A. Vitale, and J. Mendelsohn, Technometrics **11**, 265 (1969).
[153] K. Levenberg, Q. Appl. Math. **2**, 164 (1944).
[154] S. M. Goldfeld, R. E. Quandt, and H. F. Trotter, Econometrica **34**, 541 (1966).
[155] B. H. Worsley and L. C. Lax, Biochim. Biophys. Acta **59**, 1 (1962).
[156] K. Burrell, Am. J. Phys. **58**, 160 (1990).
[157] J. H. Swarte, Comput. Phys. Commun. **66**, 143 (1991).
[158] P. Deuflhard and V. Apostolesku, in *Special Topics of Applied Mathematics*, edited by J. Frehse, D. Pallaschke, and U. Trottenberg (North-Holland, Amsterdam, 1980).
[159] F. Neri, S. Patane, and G. Saitta, Meas. Sci. Technol. **1**, 1007 (1990).
[160] M. Lybanon, Am. J. Phys. **52**, 22 (1984).
[161] M. Lybanon, Comput. Geosci. **11**, 501 (1985).
[162] H. O. Hartley, Technometrics **3**, 269 (1961).
[163] I. W. Lee, M. C. Kim, and A. R. Robinson, Comput. Struct. **63**, 61 (1997).
[164] H. Späth, Commun. ACM **10**, 87 (1967).
[165] S. Oberländer, ZAMM **43**, 493 (1963) (in German).
[166] M. Berman, E. Shahn, and M. F. Weiss, Biophys. J. **2**, 275 (1962).
[167] A. Grinvald, Anal. Biochem. **75**, 260 (1976).
[168] J. Morimoto, T. Kida, Y. Miki, and T. Miyakawa, Appl. Phys. A: Solids Surf. **39**, 197 (1986).
[169] J. Morimoto, S. Tashiro, K. Tahira, and T. Miyakawa, Jpn. J. Appl. Phys., Part 1 **28**, 649 (1989).
[170] K. Tahira, M. Fudamoto, M. Tsuboyama, H. Nakashima, and T. Miyakawa, Jpn. J. Appl. Phys., Part 1 **29**, 2026 (1990).
[171] I. Isenberg, Biophys. J. **43**, 141 (1983).
[172] G. E. Bromage, Comput. Phys. Commun. **30**, 229 (1983).
[173] N. J. Clayden and B. D. Hesler, J. Mater. Res. **98**, 271 (1992).
[174] T. Awaya, Nucl. Instrum. Methods **174**, 237 (1980).
[175] T. Awaya, Nucl. Instrum. Methods **165**, 317 (1979).
[176] R. W. Hamming, *Numerical Methods for Scientists and Engineers* (McGraw-Hill, New York, 1962), p. 338.
[177] J. D. P. Donnelly, in *Methods of Numerical Approximation*, edited by D. C. Handscomb (Pergamon, New York, 1966), Chap. 15, Sec. 3.
[178] F. B. Hildebrand, *Introduction to Numerical Analysis* (McGraw-Hill, New York, 1956), Sec. 9.4.
[179] S. L. Marple, Jr., *Digital Spectral Analysis with Applications* (Prentice Hall, Englewood Cliffs, NJ, 1987).
[180] J. Myhill, G. P. Wadsworth, and G. L. Brownell, Biophys. J. **5**, 89 (1965).
[181] J. Enderlein and R. Erdmann, Opt. Commun. **134**, 371 (1997).
[182] V. V. Apanasovich and E. G. Novikov, Rev. Sci. Instrum. **67**, 48 (1996).
[183] A. C. Grove, *An Introduction to the Laplace Transform and the z Transform* (Prentice Hall, New York, 1991).
[184] M. W. Simpson-Morgan, Math. Biosci. **5**, 195 (1969).
[185] T. Sun, Z. Y. Zhang, K. T. V. Grattan, and A. W. Palmer, Rev. Sci. Instrum. **69**, 146 (1998).
[186] R. Kumaresan, D. Tufts, and L. Scharft, Proc. IEEE **72**, 230 (1984).
[187] J. N. Holt and R. J. Antill, Math. Biosci. **36**, 319 (1977).
[188] J. Loeb, G. Cahen, Automatisme **8**, 479 (1963) (in French).
[189] J. Loeb and G. Cahen, IEEE Trans. Autom. Control. **AC-10**, 359 (1965).
[190] B. Valeur and J. Moirez, J. Chim. Phys. Phys.-Chim. Biol. **70**, 500 (1973).
[191] B. Valeur, Chem. Phys. **30**, 85 (1978).
[192] N. C. Halder and A. A. Teate, J. Phys. C **2**, 10359 (1990).
[193] K. Tittelbach-Helmrich, Meas. Sci. Technol. **4**, 1323 (1993).
[194] Z. Bay, Phys. Rev. **77**, 419 (1950).
[195] I. Isenberg and R. D. Dyson, Biophys. J. **9**, 1337 (1969).
[196] S. S. Brody, Rev. Sci. Instrum. **28**, 1021 (1957).
[197] I. Isenberg and E. W. Small, J. Chem. Phys. **77**, 2799 (1982).
[198] I. Isenberg, J. Chem. Phys. **59**, 5696 and 5708 (1973).
[199] E. W. Small and I. Isenberg, J. Chem. Phys. **66**, 3347 (1977).
[200] J. Eisenfeld and D. J. Mishelevich, J. Chem. Phys. **65**, 3384 (1976).
[201] Y. L. Luke, Q. J. Mech. Appl. Math. **17**, 91 (1964).
[202] W. Fair and Y. L. Luke, Numer. Math. **14**, 379 (1970).
[203] I. M. Longman, SIAM (Soc. Ind. Appl. Math.) J. Math. Anal. **5**, 574 (1974); **23**, 439 (1972); **24**, 429 (1973).
[204] I. M. Longman, J. Comput. Appl. Math. **1**, 17 (1975).
[205] I. M. Longman, J. Inst. Math. Appl. **4**, 320 (1968).
[206] I. M. Longman, J. Comput. Phys. **10**, 224 (1972).
[207] J. E. Akin and J. Counts, SIAM (Soc. Ind. Appl. Math.) J. Math. Anal. **17**, 1035 (1969).
[208] J. E. Akin and J. Counts, J. Appl. Mech. **36**, 420 (1969).
[209] E. Yeramian and P. Claverie, Nature (London) **326**, 169 (1987).

[210] J. Aubard, P. Levoir, A. Denis, and P. Claverie, Comput. Chem. **11**, 163 (1987).
[211] O. Perron, *Die Lehre von dem Kettenbrüchen* (Chelsea, New York, 1950), p. 420 ff. (in German).
[212] I. M. Longman, Int. J. Comput. Math. **B3**, 53 (1971).
[213] J. S. Bowers, R. K. Prud'homme, and R. S. Farinato, Comput. Chem. **16**, 249 (1992).
[214] N. J. Clayden, J. Chem. Soc., Faraday Trans. **88**, 2481 (1992).
[215] S. Boudhabhay, B. Ancian, P. Levoir, R. Dubest, and J. Aubard, Comput. Chem. **16**, 271 (1992).
[216] I. M. Gelfand, G. E. Shilov, *Generalized Functions* (Academic, New York, 1964).
[217] D. Ter Haar, J. Polym. Sci. **6**, 247 (1951).
[218] R. A. Schapery, Proceedings of the 4th U.S. National Congress on Applied Mechanics (ASME, Metals Park, OH, 1962), Vol. 2, p. 1075.
[219] V. Zakian, Electron. Lett. **5**, 120 (1969).
[220] V. Zakian, Electron. Lett. **6**, 677 (1970).
[221] B. Davies and B. Martin, J. Comput. Phys. **33**, 1 (1979).
[222] D. D. Nolte and E. E. Haller, J. Appl. Phys. **62**, 900 (1987); J. Appl. Phys. **63**, 592(E) (1988).
[223] H. Stehfest, Comm. ACM **13**, 47 (1970).
[224] D. P. Gaver, Oper. Res. **3**, 444 (1966).
[225] E. L. Post, Trans. Am. Math. Soc. **32**, 723 (1930).
[226] D. Widder, Duke Math. J. **1**, 126 (1935).
[227] H. Okushi and Y. Tokumaru, Jpn. J. Appl. Phys., Part 1 **19**, L335 (1980).
[228] H. Okushi, Philos. Mag. B **52**, 33 (1985).
[229] H. Okushi, Y. Tokumaru, S. Yamasaki, H. Oheda, and K. Tanaka, Jpn. J. Appl. Phys., Part 2 **20**, L549 (1981).
[230] H. Tomokage and T. Miyamoto, Jpn. J. Appl. Phys., Part 1 **24**, 381 (1985).
[231] T. Ishikawa, Y. K. Kwon, and H. Kuwano, Appl. Phys. Lett. **47**, 1097 (1985).
[232] A. A. Istratov, O. F. Vyvenko, H. Hieslmair, and E. R. Weber, Meas. Sci. Technol. **9**, 477 (1998).
[233] A. A. Istratov, J. Appl. Phys. **82**, 2965 (1997).
[234] A. A. Istratov, Rev. Sci. Instrum. **68**, 3861 (1997).
[235] P. M. Henry, J. M. Meese, J. W. Farmer, and C. D. Lamp, J. Appl. Phys. **57**, 628 (1985).
[236] G. Ferenczi, J. Boda, and T. Pavelka, Phys. Status Solidi A **94**, K119 (1986).
[237] V. I. Turchanikov, V. S. Lysenko, and V. A. Gusev, Phys. Status Solidi A **95**, 283 (1986).
[238] L. Stuchlikova, L. Harmatha, V. Nagl, and M. Gazi, Phys. Status Solidi A **138**, 241 (1993).
[239] A. Papoulis, Q. Appl. Math. **14**, 405 (1957).
[240] H. Bateman, Duke Math. J. **2**, 569 (1936).
[241] A. Erdélyi, Philos. Mag. **34**, 533 (1943).
[242] W. T. Weeks, J. ACM **13**, 419 (1966).
[243] R. Migneron and K. S. S. Narayanan, Comput. Phys. Commun. **49**, 457 (1988).
[244] J. Shohat, Duke Math. J. **6**, 615 (1940).
[245] A. Al-Shuaibi, Inverse Probl. **14**, 1 (1998).
[246] J. W. Head, Proc. Cambridge Philos. Soc. **52**, 640 (1956).
[247] R. Piessens, J. Inst. Math. Appl. **10**, 185 (1972).
[248] W. A. Essah and L. M. Delves, Inverse Probl. **4**, 705 (1988).
[249] R. A. Spinelli, SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **3**, 636 (1966).
[250] L. M. Delves and L. F. A. El Al, Comput. J. (UK) **20**, 374 (1977).
[251] M. K. Miller and W. T. Guy, SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **3**, 624 (1966).
[252] B. S. Berger, J. Appl. Mech. **35**, 837 (1968).
[253] H. L. Armstrong, IRE Trans. Circ. Theory **CT-4**, 286 (1957).
[254] H. L. Armstrong, IRE Trans. Circ. Theory **CT-6**, 351 (1959).
[255] H. Dohler and F. Gotze, Exp. Tech. Phys. (Berlin) **35**, 401 (1987).
[256] R. Kress, *Linear Integral Equations* (Springer, Berlin, 1989), p. 232 ff.
[257] D. G. Gardner, Ann. (N.Y.) Acad. Sci. **108**, 195 (1963).
[258] M. R. Smith and S. Cohn-Sfetcu, Nucl. Instrum. Methods **114**, 171 (1974).
[259] J. W. Cooley, P. A. W. Lewis, and P. D. Welch, IEEE Trans. Audio Electroacoust. **AU-15**, 79 (1967).
[260] H. F. Silverman and A. E. Pearson, IEEE Trans. Audio Electroacoust. **AU-21**, 112 (1973).
[261] J. Schlesinger, Nucl. Instrum. Methods **106**, 503 (1973).
[262] J. W. Cooley and J. W. Tukey, Math. Comput. **19**, 297 (1965).
[263] V. I. Zubkov, Kim Kha En, A. A. Kopylov, and A. V. Solomonov, Sov. Phys. Semicond. **25**, 1302 (1991).
[264] B. Kober, W. Gruhle, and M. Hillen, Nucl. Instrum. Methods **174**, 173 (1980).
[265] S. M. Pizer, A. B. Ashare, A. B. Callahan, and G. L. Brownell, in *Concepts and Models of Biomathematics*, edited by F. Heinmets (Marcel Dekker, New York, 1969).
[266] M. R. Smith and S. T. Nichols, Nucl. Instrum. Methods Phys. Rev. **205**, 479 (1983).
[267] T. K. Linn and J. E. Dutt, Math. Biosci. **20**, 381 (1974).
[268] M. Iqbal, J. Comput. Appl. Math. **59**, 145 (1995).
[269] M. Iqbal, Comput. Phys. Commun. **88**, 43 (1995).
[270] A. Al-Shuaibi, Inverse Probl. **13**, 1153 (1997).
[271] T. L. Hayden, D. W. A. Bourne, and Y. T. Fu, Math. Biosci. **63**, 71 (1983).
[272] Y. Nakayama, S. Akita, and T. Kawamura, Jpn. J. Appl. Phys., Part 2 **27**, L320 (1988).
[273] A. Le Bloa, D. T. Quan, and Z. Guennouni, Meas. Sci. Technol. **4**, 325 (1993).
[274] H.-J. Stockmann, Nucl. Instrum. Methods **150**, 273 (1978).
[275] A. N. Tikhonov, Sov. Math. Dokl. **4**, 1035 (1963); **4**, 1624 (1963); **5**, 835 (1964).
[276] V. A. Morozov, *Regularization Methods for Ill-posed Problems* (Chemical Rubber Corp., Boca Raton, FL, 1993).
[277] V. A. Morozov, *Methods for Solving Incorrectly Posed Problems* (Springer, Berlin, 1984).
[278] A. N. Tikhonov and V. B. Glasko, Zh. Vychisl. Mat. Mat. Fiz. **4**, 564 (1964).
[279] G. Wahba, SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **14**, 651 (1977).
[280] J. Honerkamp and J. Weese, Continuum Mech. Thermodyn. **2**, 17 (1990).
[281] M. Hanke and T. Raus, SIAM (Soc. Ind. Appl. Math) J. Sci. Comput. **17**, 956 (1996).
[282] A. R. Davies and R. S. Anderssen, J. Aust. Math. Soc. B, Appl. Math. **28**, 114 (1986).
[283] Q.-N. Jin and Z.-Y. Hou, Inverse Probl. **13**, 815 (1997).
[284] P. C. Hansen, SIAM (Soc. Ind. Appl. Math.) Rev. **34**, 561 (1992).
[285] T. Reginska, SIAM (Soc. Ind. Appl. Math) J. Sci. Comput. **17**, 740 (1996).
[286] P. C. Hansen and D. P. O'Leary, SIAM (Soc. Ind. Appl. Math) J. Sci. Comput. **14**, 1487 (1993).
[287] A. R. Davies, in *Inverse Problems in Scattering and Imaging*, edited by M. Bertero and E. R. Pike (Adam Hilger, Bristol, 1992), p. 393.
[288] C. W. Groetsch, *The Theory of Tikhonov Regularization for the Fredholm Equations of the First Kind* (Pitman, London, 1984).
[289] A. B. Bakushinsky, A. V. Goncharsky, V. V. Stepanov, and A. G. Jagola, *Regularization Algorithms and A Priori Information* (MGU, Moscow, 1989) (in Russian).
[290] A. Neubauer, Inverse Probl. **5**, 541 (1989).
[291] H. W. Engl, K. Kunisch, and A. Neubauer, Inverse Probl. **5**, 523 (1989).
[292] G. Chavent and K. Kunisch, Inverse Probl. **10**, 63 (1994).
[293] O. Scherzer, Numer. Math. **66**, 259 (1993).
[294] M. Bertero, C. De Mol, and G. A. Viano, in *Inverse Scattering Problems in Optics*, edited by H. P. Baltes (Springer, Berlin, 1980).
[295] S. W. Provencher, Comput. Phys. Commun. **27**, 229 (1982).
[296] J. Morimoto, M. Fudamoto, K. Tahira, T. Kida, S. Kato, and T. Miyakawa, Jpn. J. Appl. Phys., Part 1 **26**, 1634 (1987).
[297] J. Morimoto, M. Fudamoto, S. Tashiro, M. Arai, T. Miyakawa, and R. H. Bube, Jpn. J. Appl. Phys., Part 1 **27**, 2256 (1988).
[298] M. Fudamoto, K. Tahira, J. Morimoto, and T. Miyakawa, Jpn. J. Appl. Phys., Part 1 **27**, 738 (1988).
[299] K. Tahira, J. Morimoto, and T. Miyakawa, Jpn. J. Appl. Phys., Part 1 **27**, 556 (1988).
[300] J. Yoshino, K. Tanaka, Y. Okamoto, J. Morimoto, and T. Miyakawa, Jpn. J. Appl. Phys., Part 1 **33**, 3480 (1994).
[301] D. Maier, P. Hug, M. Fiederie, C. Eiche, D. Ebling, and J. Weese, J. Appl. Phys. **77**, 3851 (1995).
[302] D. A. Batovski and Ch. M. Hardalov, J. Appl. Phys. **74**, 291 (1993).
[303] D. Batovski and Ch. Hardalov, J. Appl. Phys. **78**, 1808 (1995).
[304] L. Dobaczewski, P. Kaczor, M. Missous, A. R. Peaker, and Z. R. Zytkiewicz, Phys. Rev. Lett. **68**, 2508 (1992).
[305] L. Dobaczewski, P. Kaczor, M. Missous, A. R. Peaker, and Z. R. Zytkiewicz, J. Appl. Phys. **78**, 2468 (1995).
[306] J. Weese, Comput. Phys. Commun. **69**, 99 (1992).

[307] C. Eiche, D. Maier, M. Schneider, D. Sinerius, J. Weese, K. W. Benz, and J. Honerkamp, J. Phys. C **4**, 6131 (1992).
[308] J. Weese, Comput. Phys. Commun. **77**, 429 (1993).
[309] H. J. J. te Riele, Comput. Phys. Commun. **36**, 423 (1985).
[310] Comput. Phys. Commun., see application form in each issue of the journal.
[311] http://www.mpibpc.gwdg.de/~sproven/
[312] M. I. Tsema, Otbor Obrabotka Informatsii **84**, 99 (1992) (in Russian).
[313] V. I. Tarasov, Optoelectronics, Instrumentation Data Processing, **2**, 117 (1992).
[314] S. J. Wernecke and L. R. D'Addario, IEEE Trans. Comput. **C-26**, 351 (1977).
[315] E. T. Jaynes, Phys. Rev. **106**, 620 (1957).
[316] E. T. Jaynes, in *Maximum Entropy and Bayesian Methods*, edited by W. T. Grandy, Jr. and L. H. Schick (Kluwer Academic, Dordrecht, 1991).
[317] E. T. Jaynes, IEEE Trans. Syst. Sci. Cybern. **SSC-4**, 227 (1968).
[318] E. T. Jaynes, *Papers on Probability, Statistics and Statistical Physics* (Reidel, Dordrecht, Holland, 1983).
[319] E. T. Jaynes, in *The Maximum Entropy Formalism*, edited by R. D. Levine, M. Tribus (Massachusetts Institute of Technology Press, Cambridge, MA, 1979), p. 15.
[320] E. T. Jaynes, IEEE Trans. Syst. Sci. Cybern. **SSC-4**, 227 (1968).
[321] E. T. Jaynes, Phys. Rev. **108**, 171 (1957).
[322] J. Skilling, in *Maximum-Entropy and Baysean Methods in Science and Engineering*, Foundations, Vol. 1 (Kluwer Academic, Dordrecht, 1988), p. 173.
[323] R. M. Gray, *Entropy and Information Theory* (Springer, New York, 1990).
[324] J. Aczel, Z. Daroczy, *On Measures of Information and their Characterizations* (Academic, New York, 1975).
[325] P. M. Woodward, *Probability and Information Theory with Application to Radar* (Pergamon, Oxford, 1964).
[326] A. Mohammad-Djafari, in *Maximum Entropy and Bayesian Methods*, edited by A. Mohammad-Djafari and G. Demoment (Kluwer Academic, Dordrecht, 1993).
[327] A. Mohammad-Djafari, in *Maximum Entropy and Bayesian Methods*, edited by C. R. Smith, G. J. Erickson, and P. O. Neudorfer (Kluwer Academic, Dordrecht, 1992).
[328] U. Amato and W. Hughes, Inverse Probl. **7**, 793 (1991).
[329] S. V. Meshkov and D. V. Berkov, Int. J. Mod. Phys. C **5**, 987 (1994).
[330] P. J. Steinbach, Biophys. J. **70**, 1521 (1996).
[331] S. D. Böhmig, M. Schmid, and H. Störi, Surf. Sci. **313**, 6 (1994).
[332] G. Landl and R. S. Anderssen, Inverse Probl. **12**, 35 (1996).
[333] S. F. Gull and J. Skilling, IEE Proc. F, Commun. Radar Signal Process. **131**, 646 (1984).
[334] R. Nityananda and R. Narayan, J. Astrophys. Astron. **3**, 419 (1982).
[335] H. Gzyl, *The Method of Maximum Entropy* (World Scientific, Singapore, 1995).
[336] A. E. McKinnon, A. G. Szabo, and D. R. Miller, J. Phys. Chem. **81**, 1564 (1977).
[337] D. V. O'Connor, W. R. Ware, and J. C. Andre, J. Phys. Chem. **83**, 1333 (1979).
[338] W. A. Doolittle and A. Rohatgi, J. Appl. Phys. **75**, 4570 (1994).
[339] W. A. Doolittle and A. Rohatgi, J. Appl. Phys. **75**, 4560 (1994).
[340] R. G. Cornell, Biometrics **18**, 104 (1962).
[341] M. Piana and M. Bertero, J. Opt. Soc. Am. A **13**, 1516 (1996).
[342] G. F. Miller, in *Numerical Solution of Integral Equations*, edited by L. M. Delves and J. Walsh (Clarendon, Oxford, 1974).
[343] M. S. Mazzola, N. H. Younan, R. Soundararajan, and S. E. Saddow, Rev. Sci. Instrum. **69**, 2459 (1998).
[344] C. W. Wang and C. H. Wu, Solid-State Electron. **35**, 1771 (1992).