LBL-22206
c.2

LBL-22206
Preprint

RECEIVED
LAWRENCE
BERKELEY LABORATORY

MAR 2 1987

LIBRARY AND
DOCUMENTS SECTION

Submitted to Journal of
Applied Physics

OPTIMIZATION OF THE ENERGY RESOLUTION OF DEEP LEVEL TRANSIENT SPECTROSCOPY

D.D. Nolte and E.E. Haller

December 1986

TWO-WEEK LOAN COPY

This is a Library Circulating Copy which may be borrowed for two weeks.

Lawrence Berkeley Laboratory
University of California
Berkeley, California 94720

Prepared for the U.S. Department of Energy
under Contract DE-AC03-76SF00098

Center for Advanced Materials

LBL-22206
c.2

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

Optimization of the Energy Resolution of Deep Level Transient Spectroscopy

D. D. Nolte
Lawrence Berkeley laboratory and
Department of Physics
University of California, Berkeley 94720

E. E. Haller
Lawrence Berkeley Laboratory
and Department of Materials Science and Mineral Engineering
University of California, Berkeley 94720

Abstract

We address the problem of optimizing the energy resolution of Deep Level Transient Spectroscopy (DLTS), emphasizing techniques which are applicable for signals from defect levels which are either alloy-broadened or stress-broadened and which therefore produce non-exponential or multiexponential transients. We discuss the possibility of using a numerical inverse Laplace transform for which there is, in principle, no limit to the resolution. The sensitivity of the inversion algorithm to noise and to round-off error, however, practically limits the energy resolution to several percent of the deep level defect binding energy. This resolution is an improvement by about a factor of two over the best resolution possible from the standard rate window techniques for which there is a finite lower bound to the temperature resolution which depends on the capture cross-section of the defect and the effective time constant of the rate window, but which is independent of the defect energy.

I. Introduction

Defects in semiconductors are the origin of the electrical activity which provides the basis for integrated circuit technology. However only a few of the large variety of possible crystal defects are advantageous and can be introduced into the crystals in a controlled manner. The vast majority of defects are detrimental to device performance. Controlling or eliminating defects is difficult to achieve when the identities and structures of the defects are unknown. Because of this, considerable research is carried out characterizing the properties and origins of defects. The most important distinguishing property of a defect is the energy position of the defect in the band gap of the semiconductor. The ionization energies of shallow levels, with energies near the band extrema, can be characterized by optical excitation with high precision. Defects which have energy levels deep in the band gap, on the other hand, are not as easily characterized by optical techniques. For these defects deep level transient spectroscopy (DLTS)[1,2] provides a technique which has high sensitivity and can provide activation energies with good accuracy.

Deep level transient spectroscopy is a thermal ionization spectroscopy. The ionization signal is obtained by perturbing the occupation of defects in a reverse-biased diode, then allowing the occupation to relax to thermal equilibrium. For a single, isolated defect, the relaxation signal follows a single exponential decay in time. The time constant of the decay depends on both the bulk characteristics of the material and on the specific properties of the defect. The time constant also depends exponentially on temperature. The temperature of

the sample is used as the variable in this exponential spectroscopy, and a DLTS spectrum is obtained by ramping the temperature of the sample. Many varieties of transient spectroscopies have been developed which measure capacitive[2], current[3], or charge[4] transients in the semiconductor diode and these systems have been optimized with respect to noise[5], spatial resolution[6], and accuracy of the energy determination[7]. Despite the fact that DLTS is a thermal spectroscopy which relies on probabilistic thermal averages rather than the absolute energy differences observed in optical transitions, the energy determinations can be quite accurate when temperature and electric field effects are taken into account. However the broad response linewidth of standard DLTS as a function of temperature severely restricts the energy resolution of the technique when defect levels are closely spaced in energy. Often defects are not well isolated in their energy positions in the bandgap, rather many defects may be present which have only slightly different energies. Defect levels can be split by homogeneous perturbations, such as stress, resulting in closely spaced energy levels, or they can be perturbed inhomogeneously, resulting in continuous distributions of energies. In these cases the DLTS signal is composed of multiple exponentials. The problem of DLTS resolution therefore is to separate the individual components as accurately as possible.

The problem of decomposing a multiple exponential into its constituent parts is one of the oldest and yet most persistent problems of functional analysis. In 1795 Prony[8] devised an algebraic scheme which could separate a small number of exponentials with similar amplitudes but substantially different time constants. The method does not differentiate well between exponentials with similar time constants

or with widely different amplitudes. Prony recognized this limitation as fundamental, and this view has been echoed by many modern authors[9,10,11]. Currently, roughly four classes of approach to the problem of deconvoluting a multiple exponential can be identified[12]: curve peeling methods[13], which graphically subtract long lived decays from short lived components; least squares, both linear and non-linear[14,15,16]; algebraic methods, also known as Prony methods[8,17,18]; and transform methods, which are deconvolution methods[19,20,21,22]. Many of these techniques have been reviewed and compared by Thomasson et. al.[23]. Recently a powerful analytic technique has been developed called Pade-Laplace analysis[24] which combines the transform and algebraic methods into the most promising technique yet for decomposing a sum of discrete exponentials. Interestingly, Pade-Laplace analysis reproduces the two-century old Prony method as a limiting case. However Pade-Laplace analysis is severely limited when the distribution of exponentials is continuous because the method actually locates and counts the number of poles in the complex plane resulting from a Laplace transform of the signal. The counting method accumulates larger errors as the number of poles increases, making the technique tedious and inaccurate for continuous distributions. It may be possible to optimize the Pade-Laplace method to be workable with continuous distributions, however we do not undertake such an analysis in this paper.

An important distinction can be made between those methods which make many or strong assumptions about the form of the signal, and those which make no or few assumptions about the form of the signal. The former are essentially fitting routines, in which the form of

the solution is assumed, and free parameters are altered to yield agreement between the signal and the assumed form. The latter are more general and spectroscopic in nature. The transform methods fall into this latter class.

The fitting methods can be quite accurate when the constituent signals are true exponentials, and when the number of constituents is known. However these methods lack the generality of a true spectroscopic approach, such as Fourier transform spectroscopy, which makes no assumptions about the signal other than that it is composed of a superposition of sinusoids. Often the number of components in a signal is the very information which is being investigated. In the case of uniaxial stress DLTS[25,26,27] the number of peaks and intensity ratios that result after the application of uniaxial stress gives an indication of the defect symmetry. The interpretation of the data would be more straightforward if these parameters were the result of a standard set of general operations rather than from assumptions in a fitting routine. In addition, situations can occur for which the fitting methods will not apply. One of these situations arises in the routine DLTS of ternary III- V compounds. The DLTS transients are found to be non-exponential, stemming from alloy broadening of the DLTS peaks[28]. A similar situation arises in uniaxial stress DLTS; the induced strains in the DLTS stress sample may be inhomogeneous, leading to non-exponential transients of complicated form which cannot be fit.

The transform methods, on the other hand, are general and provide the basis for a true spectroscopy of continuous distributions of exponential transients. The rate window technique of Lang[2] as well as the inverse Laplace transform belong to this class of transform techniques. We begin with the inverse Laplace transform in section II to point out the

strengths and drawbacks of the inverse Laplace transform as applied to exponentials. In section III we discuss, in depth, the fundamental lower limit to the resolution of the rate window technique.

II. Inverse Laplace Transform

The goal of DLTS signal processing is to convert the raw deep level capacitive or current transient into a system response whose variation as a function of temperature gives the thermal emission rate, the emission cross section, and the concentration of each deep level in the band gap. The activation energy and capture cross section comprise the signature of the deep level. The problem of DLTS resolution therefore is to optimally separate signatures of two or more levels which have comparable emission rates. Assuming the the raw DLTS signal consists of a sum of exponential decays,

$$S(t) = \sum_{i}^{n} C_i e^{-t/\tau_i} \quad , \qquad (1)$$

the problem is solved in principle by taking the inverse Laplace transform of the raw signal. Because exponentials and delta functions are Laplace transform pairs, the result of the inverse Laplace transform would be a sum of delta functions,

$$R(s) = \sum_{i}^{n} C_i \delta(s - 1/\tau_i(T)) \qquad (2),$$

where the τ_i are the emission time constants as a function of temperature.

This procedure appears to be analogous to the use of Fourier transforms in Fourier Transform Spectroscopy (FTS), which is based on

sinusoids and delta functions as Fourier transform pairs. FTS relies on the orthogonality of sinusoids to project the contribution from each frequency out of the signal. In practice the transform integration cannot be taken to infinity, but must be terminated at some limit, L. This finite integration contributes a line width to the response function which varies as 1/L. Because the signal in FTS is composed of periodic functions whose amplitudes remain constant, the integration can be carried over many periods, thereby increasing the signal to noise and decreasing the line width. This allows for excellent frequency and therefore energy resolution in Fourier Transform Spectroscopy.

Unlike sinusoids, however, exponentials are not orthogonal along the real axis. The contribution of each exponential to the signal cannot be projected out, as in FTS, by taking an inner product defined as an integral along the real axis. This is reflected by the fact that the inverse Laplace transform requires a contour integration in the complex plane. Performing the integration numerically in this case would require either a knowledge of the signal function S(t) throughout the the complex plane, or a knowledge of its poles in the complex plane. Yet from experimental observation only the values of the signal function along the positive real axis are known. One possible approach might be to analytically continue the signal function into the complex plane and find the complex poles. However this is a cumbersome, computationally intensive process which would be difficult to implement in real time data acquisition.

Barring knowledge of the analytic structure of the signal in the complex plane, statistical arguments can be invoked to develop efficient algorithms that can numerically invert Laplace transforms. No single well-behaved algorithm for numerical inversion of the Laplace transform can be defined, however, because the inverse Laplace operator is an

unbounded operator; inverse functions are unstable with respect to small changes in the direct function, that is, arbitrarily small changes in the direct function can cause arbitrarily large changes in the inverse function. Nonetheless, an accurate algorithm has been developed by Stehfest[29] based on a statistical expectation function defined by Gaver[30]. Given a real function S(t), the algorithm calculates an approximation of the inverse R(s) at s

$$R(s) = \frac{\ln 2}{s} \sum_{i=1}^{N} V_i \, S(i \ln 2/s) \qquad (3)$$

where

$$V_i = (-1)^{N/(2+i)} \sum_{k=[i+1/2]}^{\mathrm{Min}(i,N/2)} \frac{k^{N/(2+i)} \, (2k)!}{(N/2-k)!k!(k-1)!(2k-1)!} \qquad (4)$$

This algorithm gives highly accurate results for situations in which both the transform and its inverse are smooth functions, but its convergence properties worsen dramatically as the functions vary rapidly or contain discontinuities. While a decaying exponential is quite smooth, its inverse transform, the delta function, is not. This fact from the outset indicates that there will be severe problems with convergence for this case. Convergence is improved by choosing a large N, but because of the numerous factorials in V_i, N can only be as large as the number of significant figures used in the computation, otherwise round-off errors destroy the convergence. The result of this algorithm applied to e^{-t} and to e^{-2t} with double precision arithmetic (N=24) on a VAX 780 is given in

Fig. 1. The peaks occur at s = 1 and s = 2, and have a bandwidth (in the transform space) of 50%. The height of the response peak scales linearly with the exponential prefactor, but diminishes as $1/s_i$. The response of the algorithm to two exponentials is linear, which is essential for signal analysis. The highest resolution of this algorithm for N=24 is achieved for a ratio of the time constants equal to 1.5. To couch this result in physical terms we will consider the EL6 defect in GaAs, the dominant bulk defect in that material, to be a "typical" deep level. EL6 has an energy level at E_C - .350 meV with a capture cross section $\sigma = 1.5 \times 10^{-13}$ cm^2. The DLTS transient from EL6 has a time constant of 3 ms at 200 K. For this "typical" deep level a resolved ratio of time constants equal to 1.5 corresponds to an energy resolution of 2% of the defect binding energy. This value is to be compared to the typical resolution of standard DLTS of between 9% and 15% (c.f. sect.III). Therefore in the absence of noise, the Gaver-Stehfest algorithm can provide a substantial improvement in deep level energy resolution.

When a random gaussian distribution of errors is added to the signal, the maximum allowable value of N decreases. For large values of N the algorithm amplifies the noise over the signal, while for small values of N the peak becomes broader. The optimum value of N depends on the noise variance of the signal. For a noise variance of 0.3% of the peak signal added to each data point in eq.(3), N = 6 is the largest value of N which will give acceptable results. Fig. 2 displays the responses of the algorithm to the exponential e^{-t} for two sets of errors as well as the response for no errors. For the "typical" deep level, this corresponds to an energy resolution of about 4%, which is still an improvement by about a factor of two over the best possible resolution from standard DLTS.

The Gaver-Stehfest algorithm is ideally suited for implementation on any system which can integrate and digitize the transient signal. The algorithm is especially convenient for real-time analysis because the vector V_i of eqns. (3) and (4) only needs to be calculated once at the start of the experiment. A DLTS system using the Gaver-Stehfest algorithm can be run in two modes: a scanning mode, and a high resolution mode. In the scanning mode, a response as a function of temperature can be defined by selecting a single value of s in eqn. (3) and then ramping the temperature of the DLTS diode. In this case the real-time analysis would require signal integration, data smoothing, and a single summation in eqn. (3). The data smoothing is necessary because of the sensitivity of the Gaver-Stehfest algorithm to noise. Data smoothing routines can be fast and simple and should not significantly interrupt the data acquisition. The data integration is necessary to reduce the noise and is the limiting factor on the speed of the temperature scan. In the high resolution mode the temperature of the DLTS diode can be stabilized so that significant integration time is available. After data smoothing, the full response R(s) of eqn. (3) is calculated and filtered. Because the response function is linear, the presence of multiple transients can be detected by analyzing the response lineshape in the transform space.

The implementation of the Gaver-Stehfest algorithm for DLTS systems does have drawbacks, the primary drawback being its extreme sensitivity to noise. In addition, the implementation requires analog-to-digital conversion and a computer controller. Analog systems, on the other hand, have the advantage of simplicity and good stability with respect to noise. Therefore the standard rate window technique remains a vital analysis technique for DLTS. We consider the limit of resolution for the rate window technique next.

III. Rate Window Technique

The basic ideas of the rate window technique as applied to DLTS have been discussed elsewhere[2,5,31] , but the details of the response lineshape have not. Therefore we shall investigate the effects of the various operations on the lineshape and describe some general trends and conclusions which are valid for the many types of weighting functions that are used in the rate window techniques. We find that a finite lower bound exists for the resolution which is independent of the energy of the deep level. This bound does depend on properties of the bulk material as well as on the capture cross section of the defect.

After a deep level defect in the depletion region of a diode is filled by a bias reduction or a forward bias filling pulse, it will thermally emit the captured carriers into the valence or conduction band with the rate

$$e_n = \sigma_0 N_v v_{th} e^{-\Delta H/kT} \qquad (5)$$

derived through detailed balance, where σ_0 is the capture cross-section (including an entropy term), N_v is the effective density of states at the band edge, v_{th} is the thermal velocity of carriers in the band, and ΔH is the emission activation enthalpy of the defect.

The rate window technique quite simply gives a response that peaks when the emission rate is equal to a specified value. The response is formed by applying three linear operations to the raw, measured signal.

First, a baseline time t_0 is defined and the value of the signal at this time is subtracted from the raw signal. Second, the baseline restored signal is multiplied by a weighting function with a characteristic time constant related to t_0. Third, the resultant function is integrated over time. Because the DLTS signal is composed of decaying exponentials, the second and third operations together are equivalent to performing a Laplace transform of the weighting function. The high temperature side of the response lineshape can therefore be understood by looking at the trends of the Laplace transforms as the weighting functions are varied. The baseline restoration, however, modifies the Laplace transform for slow emission rates at the lower temperatures. The low temperature side of the response lineshape is therefore dominated by the effect of baseline restoration.

A general, analytic expression can be derived in a straightforward manner for the full width half max (FWHM) linewidth of the response as a function of temperature. If e_1, T_1 and e_2, T_2 are the emission rates and temperatures at which the response is half of the maximum, and e_0, T_0 are the emission rate and temperature at which the response is a maximum, then the temperature resolution is given by

$$\frac{\Delta T}{T} = \ln(\sigma_0 N_0 v_0/e_0) \left[\frac{1}{\ln(\sigma_0 N_1 v_1/e_1)} - \frac{1}{\ln(\sigma_0 N_2 v_2/e_2)} \right] \quad (6)$$

where N_1, N_2, and N_3 are the band density of states at T_1, T_2, and T_3 respectively.

To facilitate the comparison between the various weighting functions and baseline restoration times, this equation can be rewritten as

$$\frac{\Delta T}{T} = \ln(\sigma_0 N_0 v_0 \tau^* t_0)\left[\frac{1}{\ln(\sigma_0 N_1 v_1 \tau^* t_0)+\ln(x_1)} - \frac{1}{\ln(\sigma_0 N_1 v_1 \tau^* t_0)+\ln(x_2)}\right] \qquad (7)$$

where τ^* is the effective time constant defined as $1/e_0 t_0$, and x_1 and x_2 are the ratios of the half max time constants over $\tau^* t_0$.

This equation is quite general, yet several observations can be made regarding its qualitative behavior. First, the linewidth does not depend explicitly on the binding energy of the deep level. Second, the linewidth depends only weakly on the temperature through the temperature dependence of the density of states and the thermal velocities. Third, the exact value of the linewidth depends sensitively on the cancellation between the two terms in parentheses. Typical values for $\ln(\sigma N v \tau^* t_0)$ are of the order of 20, while $\ln(x_1)$ and $\ln(x_2)$ are typically of the order of 1. Therefore the linewidth is determined predominantly through the defect capture cross section and the baseline restoration time t_0 (both of which can vary by many orders of magnitude) and depends rather weakly on the the specific choice of weighting function (which defines τ^*).

As mentioned earlier, the details of the high temperature side of the response lineshape are determined from the properties of the Laplace transform of the weighting functions. Neglecting baseline restoration, the response for a weighting function w(s,t) is

$$R(s,\tau) = s\int e^{-t/\tau}\, w(s,t)\, dt \qquad (8)$$

where s is the the weighting function rate, and τ is the emission rate of the deep level as a function of temperature. This response is just the Laplace transform of the weighting function. The transforms of three common weighting functions are given in Table I. The high temperature lineshape is determined by the numerators in Table I because the product $s\tau$ vanishes as the temperature increases above the temperature when $s\tau = 1$. Clearly, the response for the sine function vanishes faster at high temperatures than the responses for the cosine and decreasing exponential functions.

The functional form of the high temperature response of these and all other Laplace transforms is simply the Laplace transform of the leading term in the Taylor's expansion of the weighting function. Therefore if the leading term in the weighting function goes as t^n, then the high temperature dependence of the response goes as $(s\tau)^{n+1}$. To increase the steepness of the slope, which decreases the high temperature half-width at half-max of the final response, the leading term in the Taylor's expansion of the weighting function should be of as high an order as possible. This is nothing more than *weighting to later times* which reflects the fact two decaying exponentials separate hyperbolically with increasing time. The price, of course, for the decrease in linewidth is the corresponding decrease in signal to noise.

Baseline subtraction is necessary both to remove any DC drifts in the signal and to make measurements feasible in a reasonable length of time. Baseline restoration does not affect the high temperature response of the Laplace transform, so the remarks about weighting to later times is still valid. However the baseline subtraction does cut off the low temperature response of the Laplace transform, providing the

characteristic DLTS lineshape.

The time constant of the weighting function, (1/s), and the baseline restoration time, t_0, both provide characteristic measurement times. They are not equal in general, though they usually are both of the same order of magnitude. As s or t_0 or the functional form of the weighting functions are changed, the emission rate e_0 for which the response function is a maximum changes. To compare the various response functions, the dimensionless effective time constant $\tau^* = 1/e_0t_0$ is used which is related to s implicitly. As a rule of thumb $\tau^* \approx 2/st_0$, i.e. $e_0 \approx s/2$, but this is not true in general, especially when $1/s > t_0$. For a given weighting function, as s decreases the weighting function weights to later times with respect to t_0, decreasing the linewidth. This causes τ^* to increase towards unity.

Changing either s or t_0, or both, changes the temperature at which the DLTS response peaks. This artificially affects the calculated linewidth through the temperature dependence of the density of states. Therefore to compare the results for various weighting functions with different s's, and t_0's, it is necessary to "normalize" the responses such that the peak always occurs at the same temperature. This is done in principle by choosing a set function and a set ratio $1/st_0$, and then changing s and t_0 simultaneously until the peak occurs at the same temperature T_0. Since the ratio $1/st_0$ implicitly determines the effective time constant τ^* for the function, the results to be compared can all be plotted vs. τ^*.

The full width half max linewidths for nine functions covering a range of $1/st_0$ are plotted in Fig. 3 vs. τ^*. The magnitude of the linewidths depend on $\sigma N v \tau$, which is taken equal to 10^9 in this example.

Several features can be noted. First, the linewidth for a set function decreases as τ^* increases; this is nothing more than the effect of decreasing the linewidth by weighting to later times. Also, the linewidth from several different functions which have the same τ^* decreases as the order of the leading term in the Taylor's expansion of the function increases. A deviation from this last observation occurs for the sine function with low τ^*. This is due to the oscillatory nature of the sine function leading to an additional effective baseline subtraction when $st_0 > 2$. Furthermore, the weighting functions which are mononomials have only one τ^*, regardless of the choice of s, and they occur at the limits of the relevant weighting functions as s goes to zero such that only the first term in the Taylor's series of the weighting function remains. In the Appendix the values of τ^* are given for the nine weighting functions of Fig. 3 .

The most important result from Fig. 3 is that there is an absolute lower bound to the linewidth. This limit is approached as τ^* approaches unity. Of course the linewidth limit, which is always finite, is approached with vanishing signal to noise. The Taylor's expansion of $\delta(t-t_0)$ around $t = 0$ has the highest order leading term of any function defined on the interval $[0,t_0]$. Therefore the delta function produces the minimum linewidth of all weighting functions.

The resolution limit as a function of $\sigma N v t_0$ is plotted in Fig. 4 . The linewidth decreases for increasing σ and t_0. The capture cross section, σ, can vary in value from 10^{-12} cm^2 for defects with coulombic excited states, to 10^{-22} cm^2 for deep repulsive centers. Therefore semi-shallow, effective-mass-like states typically have quite narrow linewidths. Although σ is a non-variable property of the deep

level defect, t_0 is set by the experimenter, and depending on the apparatus can be increased from microseconds to seconds. The values in Fig. 4 are only the low bound on the possible linewidth. The actual linewidth will be larger than this value, the amount depending on the choice of weighting function.

As stated previously the minimum linewidth is approached only at a cost of the signal-to-noise ratio. For a single shot, the signal is

$$S = \int_0^{t_0} w(t)\, S(t)\, dt \qquad (9)$$

where $S(t) = s(t) - s(t_0)$, $s(t)$ being the raw signal.

The noise is

$$N = \left\{ \int_0^{t_0} \left| w(t) \right|^2 dt \right\}^{1/2} . \qquad (10)$$

For repetitive signal averaging the signal to noise ratio becomes

$$S/N\big|_{rep} = \sqrt{\text{rep rate}} * S/N\big|_{\text{one shot}} \qquad (11)$$

where $\sqrt{\text{rep rate}} = 1/\sqrt{t_0}$.

The signal-to-noise ratios for eight of the weighting functions in Fig. 3 are plotted in Fig. 5. The signal-to-noise ratio decreases roughly exponentially with increasing τ^*. This reflects the fact that weighting to later times decreases the signal-to-noise ratio by weighting the tails of exponentially decreasing signals. At small effective time constants the signal-to-noise ratio also falls off because the time constant of the

weighting function is considerably smaller than the measurement time, t_0, creating considerable "dead" time before the next repetition. As pointed out by Miller et. al.[5], the largest signal-to-noise ratio is obtained for the decaying exponential weighting function. The largest signal-to-noise ratio occurs when the time constant of the exponential is roughly half of t_0.

IV. Conclusion

We have investigated the efficiency of inverse Laplace transform and rate window techniques in decomposing a signal consisting of a sum of decaying exponentials into its constituent parts. These techniques are all spectroscopic transform techniques that give general results without making strong assumptions about the original form of the signal.

The rate window technique is the more stable technique with respect to noise, but has the lower resolution. An absolute lower bound on the temperature linewidth is found which is independent of the defect activation energy. This lower bound does depend on the capture cross section and on the baseline restore time, both of which can vary by many orders of magnitude. The smallest linewidths are given by using a weighting function that has a high order first term in its Taylor's expansion. This explains the empirical result that the best practical resolution is obtained with the double boxcar using narrow windows and a long delay after the pulse[7]. With these constraints the smallest achievable energy linewidth is limited to around 7%, though the most usual situations give linewidths that vary from 9% to 15%.

The Gaver-Stehfest algorithm for the inverse Laplace transform can provide a substantial improvement in resolution over the standard rate

window technique. While the algorithm is unstable with respect to noise, simple data integration and smoothing are sufficient to bring the algorithm into an efficient range. For a common signal-to-noise ratio of 333:1 an energy resolution of about 4% may be possible for "typical" deep levels. The algorithm is also well suited for real-time data analysis and should be easily implemented on computer-controlled DLTS systems.

Acknowledgements

We would like to thank Kent Owen for helpful discussions concerning inversion techniques, and Laura J. Pyrak for her help in running the inversion algorithm on the VAX 780. We would also like to thank Pat Mooney for pointing out the special relevance of DLTS resolution for alloy systems. This work was supported by the Director, Office of Energy Research, Office of Basic Energy Research, Materials Science Division of the U. S. Department of Energy under contract DE-AC03-76SF00098.

Appendix

The effective time constants for the various weighting functions must be determined numerically as the signal time constant which maximizes the response

$$R = s \int_0^{t_0} w(s,t) \, [\, S(t) - S(t_0) \,] \, dt$$

where s is the rate of the weighting function, and S(t) is the signal. The effective time constant is determined by both s and t_0 through the dimensionless parameter $n = 1/st_0$. The values for the effective time constants are:

$\delta(t-1/s)$:

n	.1	.2	.3	.4	.5	.6	.7	.8	.9
τ^*	.35	.50	.59	.66	.72	.79	.84	.90	.93

e^{-st}:

n	.1	.2	.5	1.0	2.0	5.0
τ^*	.36	.42	.49	.52	.54	.55

e^{st}:

n	.1	.2	.5	1.0	2.0	5.0
τ^*	.89	.78	.65	.60	.58	.57

sin(st):

n	.1	.2	.5	1.0
τ^*	.34	.52	.67	.69

$e^{st} - 1$:

n	.2	.5	1.0	2.0
τ^*	.80	.73	.71	.70

(1): $\tau^* = .56$

(st): $\tau^* = .69$

$(st)^2$: $\tau^* = .76$

$(st)^3$: $\tau^* = .80$

References:

[1]C. T. Sah, L. Forbes, L. L. Rosier, and A. F. Tasch Jr., Solid-state Electron. 13,759 (1970)

[2]D. V. Lang, Thermally Stimulated Relaxation in Solids, ed. , Springer-Verlag (1979) pp.109-134

[3]J. A. Borsuk and R. M. Swanson, IEEE Trans. Electr. Dev. ED-27,2217 (1980)

[4]J. W. Farmer, C. D. Lamp, and J. M. Meese, Appl. Phys. Lett. 41,1063 (1982)

[5]G. L. Miller, J. V. Ramirez, D. A. H. Robinson, J. Appl. Phys. 46, 2638 (1975)

[6]G. P. Li and K. L. Wang, J. Appl. Phys. 57, 1016 (1985)

[7]W. R. Thurber, r. A. Forman, and W. E. Phillips, J. Appl. Phys. 53,7397(1982)

[8]R. Prony, J. Ecole Polytechnique Paris 1, 24 (1795)

[9]C. Lanczos, Applied Analysis, Prentice Hall, Inc., Englewood Cliffs, N.J. 1959 pg. 287

[10]R. E. Bellman, R. E. Kalba, and J. Lockett, Numerical inversion of the Laplace Transform, American Elsevier, New York, 1966

[11]R. W. Hamming, Numerical Methods for Scientists and Engineers, McGraw Hill, New York 1962 pp.338-342

[12]J. Aubard, P. Levoir, A. P. Denis and P. Claverie (unpublished)

[13]H. D. Van Liew, Science 138, 682 (1962)

[14]D. W. Marquardt, SIAM J. 11, 431 (1963)

[15]K. Levenberg, Quart. Appl. Math. 2, 164 (1944)

[16]A. Lemaitre and J. Malenge, Comp. and Biomed. Res., 4, 555 (1971)

[17]F. B. Hildebrand, Introduction to Numerical Analysis, McGraw-Hill, New York, 1956, chap. 9, sec. 9.4:378-382

[18]J. Myhill, G.P. Wadsworth and G. L. Brownell, Biophys. J. 5, 98 (1965)

[19]D. G. Gardener, J. C. Gardener, G. Lausch and W. W. Meinke, J. Chem. Phys. 31, 978 (1959)

[20]S. W. Provencher, J. Chem. Phys. 64, 2772 (1976)

[21]S. W. Provencher, Biophys. J. 16, 27 (1976)

[22]Y. Balcou, Int. J. of Modelling and Simulation 1, 47 (1981)

[23]W. M. Thomasson and J. W. Clark, Jr., Math. Biosc. 22, 179 (1974)

[24]P. Claverie and A. Denis, (unpublished)

[25]J. M. Meese, J. W. Farmer, and C. D. Lamp, Phys. Rev. Lett. 51, 1286 (1983)

[26]D. D. Nolte, E. E. Haller, and P. Omling, Proc. 14th Int. Conf. on Defects in Semiconductors (1986)

[27]J. L. Benton, K. M. Lee, P. E. Freeland, and L. C. Kimerling, Proc. 13th Int. Conf. on Defects in Semiconductors, AIME 14a,647 (1984)

[28]P. Omling, L. Samuelson and H. G. Grimmeiss, J. Appl. Phys. 54, (1983)

[29]H. Stehfest, Comm. of the ACM 13, 47 (1970)

[30]D. P. Gaver, Oper. Res. 14, 3, 444 (1966)

[31]G. L. Miller, D. V. Lang, L. C. Kimerling, Ann. Rev. of Mat. Sci., (Annual Reviews, Inc.) 1977, pp. 377-448

Table I.

$$R(s,\tau) = s \int_0^\infty e^{-t/\tau}\, w(s,t)\, dt$$

$$1/\tau = \sigma_0 N v e^{-\Delta H/kT}$$

Laplace: $w(s,t) = e^{-st}$

$$R(s,\tau) = \frac{s\tau}{1 + s\tau}$$

Sine: $w(s,t) = \sin st$

$$R(s,\tau) = \frac{(s\tau)^2}{1 + (s\tau)^2}$$

Cosine: $w(s,t) = \cos st$

$$R(s,\tau) = \frac{s\tau}{1 + (s\tau)^2}$$

Figure Captions:

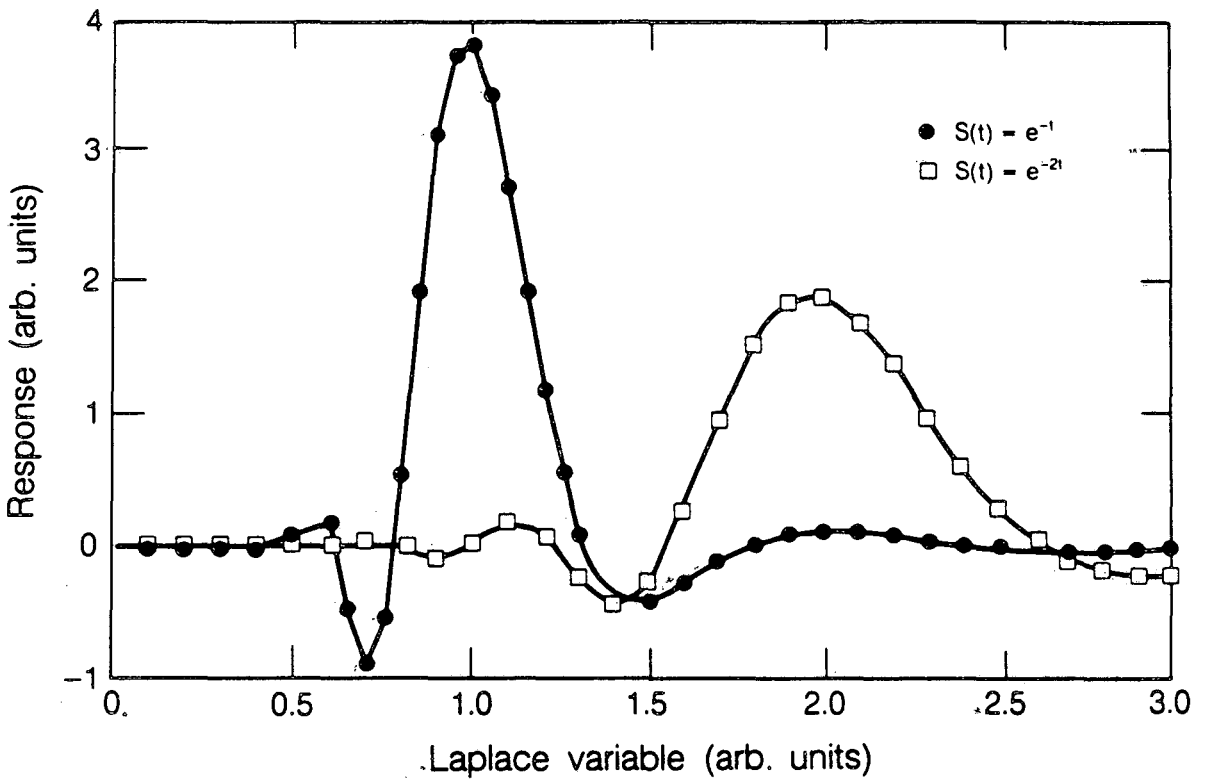
Fig. 1 : Result of the Gaver-Stehfest algorithm for the inverse Laplace transform performed on $S(t) = e^{-t}$ and $S(t) = e^{-2t}$ using N = 24.

Fig. 2 : Result of Gaver-Stehfest inverse Laplace tranform algorithm for N = 6 and a noise variance of 0.003 per point in the signal function. The data points are the results for two data sets with pseudo-random noise added to $S(t) = e^{-t}$. The full curve is the result of the algorithm for no noise.

Fig. 3 : The full-width-half-max of the DLTS response as a function of the effective time constant τ^* for nine different weighting functions. The values are computed assuming $\sigma N v t_0 = 10^9$. The functions 1, (st), $(st)^2$, and $(st)^3$ have a uniquely defined effective time constant independent of s which are the limit points of the Taylor's expansions of the other weighting functions. All the other functions have effective time constants that change as s is changed. The values of τ^* for different s are given in the Appendix.

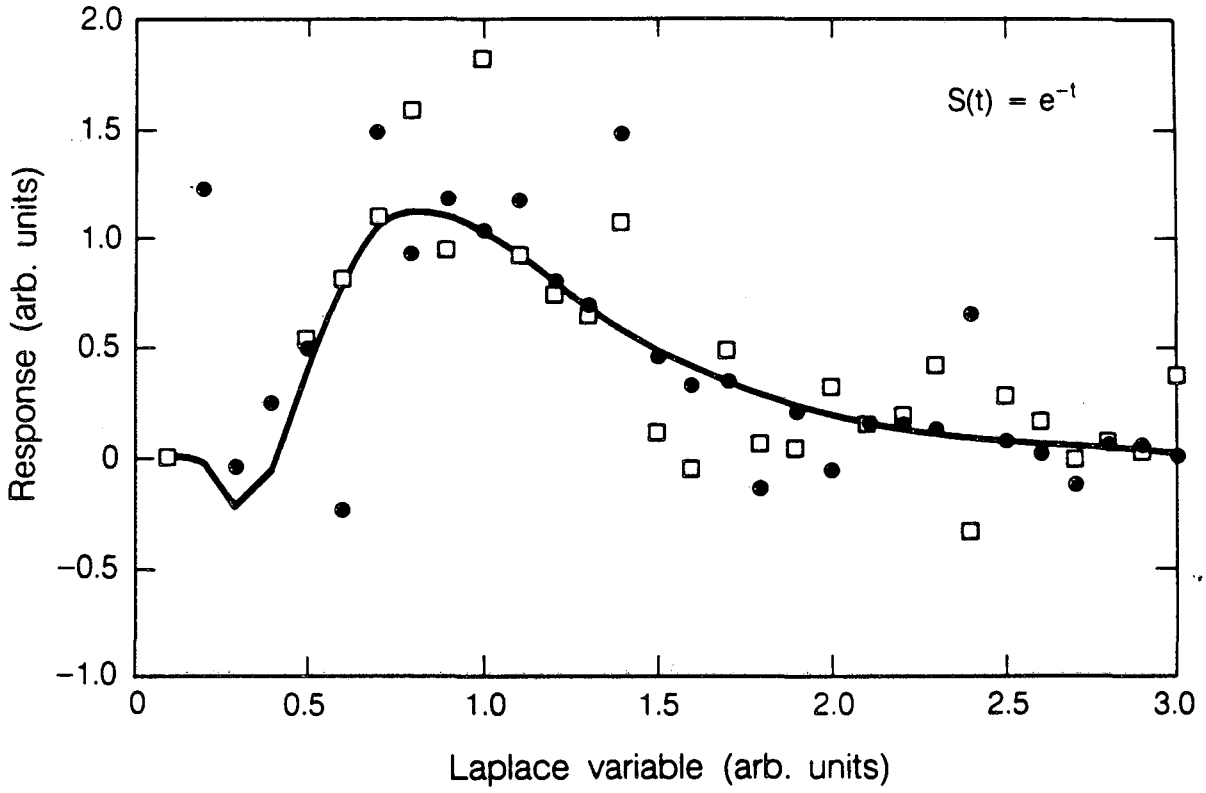
Fig. 4 : The absolute lower bound on DLTS resolution as a function of $\sigma N v t_0$. These values were extrapolated from the $\tau^*=1$ limit of the values plotted in Fig. 3 for varying $\sigma N v t_0$.

Fig. 5 : Signal to noise ratio for eight weighting functions as a function of effective time constant. The monomials are again the limit points for the other weighting functions as s -> 0.



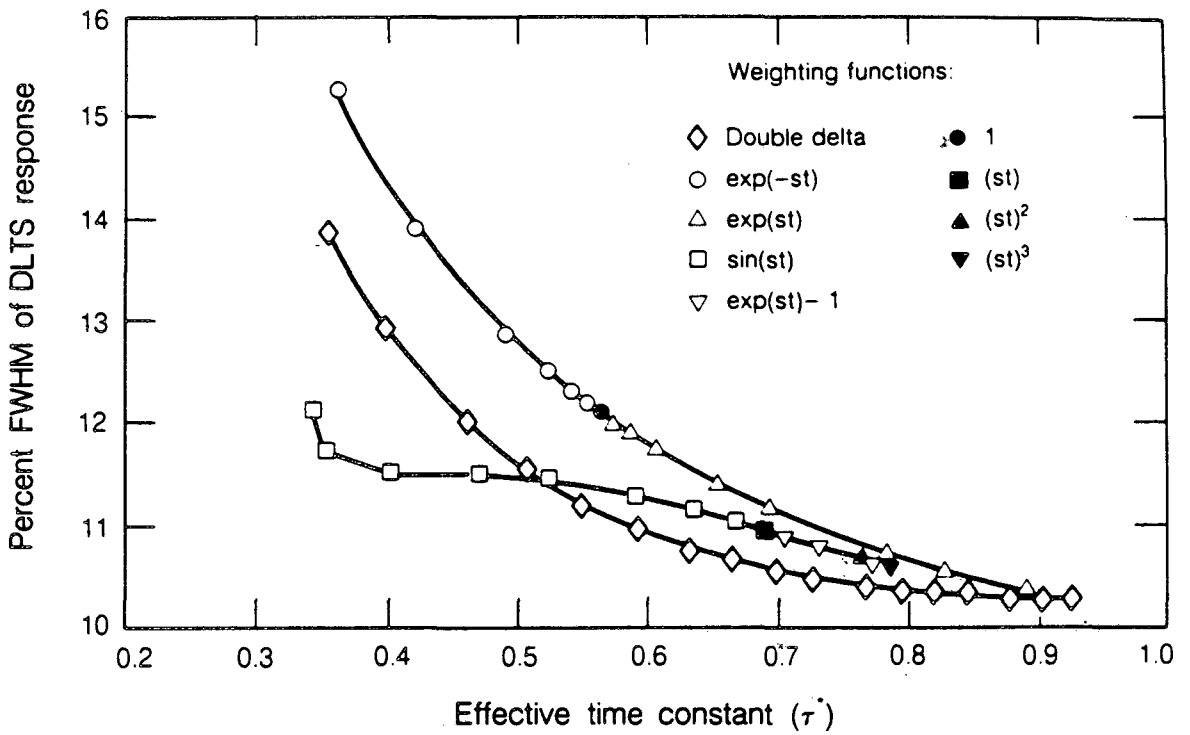
XBL 865-10118

Figure 1.



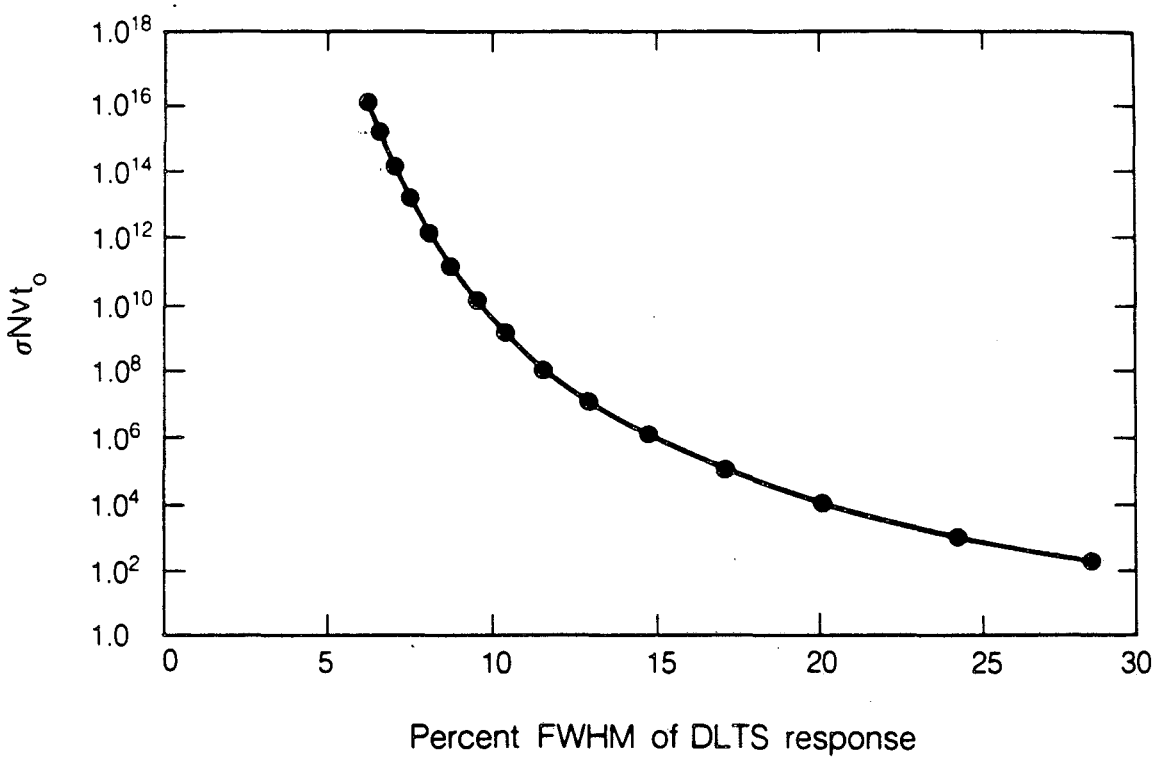
XBL 865-10119

Figure 2.



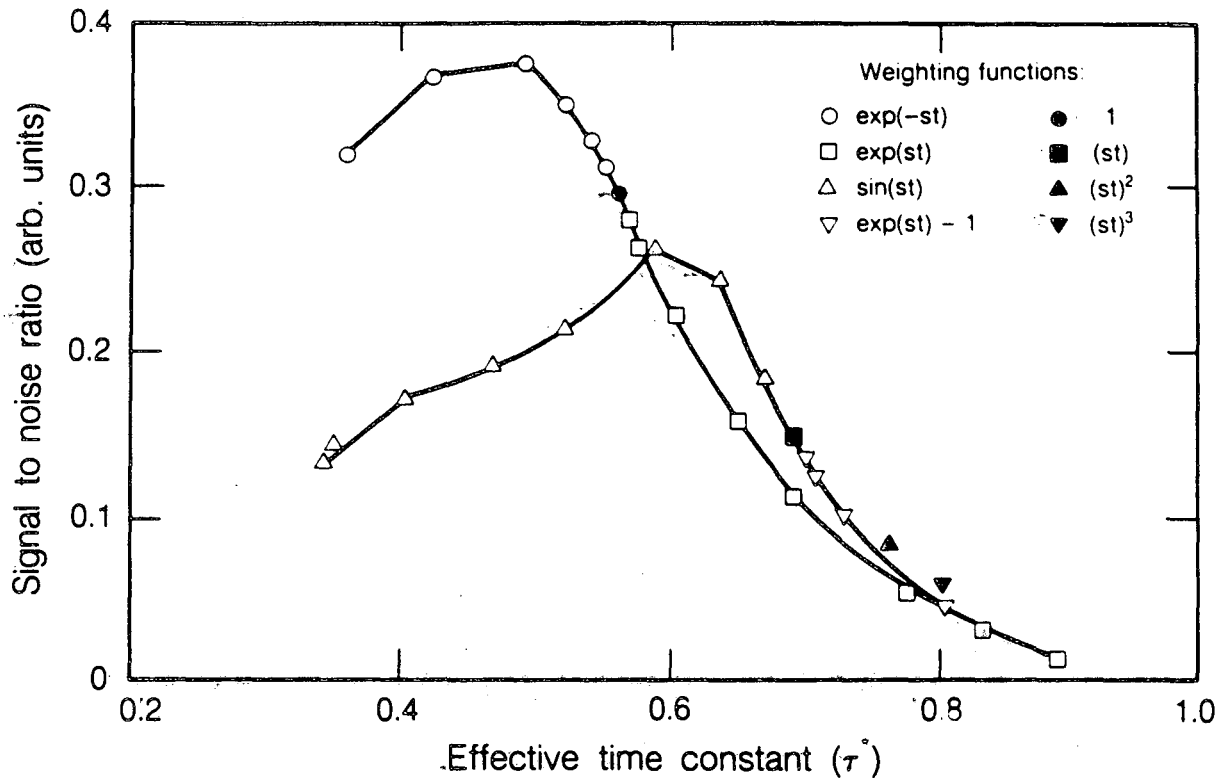
XBL 865-10121

Figure 3.



XBL 865-10122

Figure 4.



XBL 865-10123

Figure 5.

This report was done with support from the Department of Energy. Any conclusions or opinions expressed in this report represent solely those of the author(s) and not necessarily those of The Regents of the University of California, the Lawrence Berkeley Laboratory or the Department of Energy.

Reference to a company or product name does not imply approval or recommendation of the product by the University of California or the U.S. Department of Energy to the exclusion of others that may be suitable.

LAWRENCE BERKELEY LABORATORY
TECHNICAL INFORMATION DEPARTMENT
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720