MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Multi-exponential Lifetime Extraction in Time-logarithmic Scale

Knyazev, A.; Gao, Q.; Teo, K.H.

TR2016-062 July 2016

Abstract

Methods are proposed for estimating real lifetimes and corresponding coefficients from real-valued measurement data in logarithmic scale, where the data are multi-exponential, i.e. represented by linear combinations of decaying exponential functions with various lifetimes. Initial approximations of lifetimes are obtained as peaks of the first derivative of the data, where the first derivative can, e.g, be calculated in the spectral domain using the cosine Fourier transform. The coefficients corresponding to lifetimes are then estimated using the linear least squares fitting. Finally, all the coefficients and the lifetimes are optimized using the values previously obtained as initial approximations in the non-linear least squares fitting. We can fit both the data curve and its first derivative and allow simultaneous analysis of multiple curves.

Industrial Conference on Data Mining (ICDM)

Multi-exponential Lifetime Extraction in Time-logarithmic Scale*

Andrew V. Knyazev**, Qun Gao, and Koon Hoo Teo

Mitsubishi Electric Research Labs (MERL)
201 Broadway, Cambridge, MA 02139, USA

Abstract. Methods are proposed for estimating real lifetimes and corresponding coefficients from real-valued measurement data in logarithmic scale, where the data are multi-exponential, i.e. represented by linear combinations of decaying exponential functions with various lifetimes. Initial approximations of lifetimes are obtained as peaks of the first derivative of the data, where the first derivative can, e.g, be calculated in the spectral domain using the cosine Fourier transform. The coefficients corresponding to lifetimes are then estimated using the linear least squares fitting. Finally, all the coefficients and the lifetimes are optimized using the values previously obtained as initial approximations in the non-linear least squares fitting. We can fit both the data curve and its first derivative and allow simultaneous analysis of multiple curves.

Keywords: lifetime extraction, exponential, multi-exponential, least squares, fitting, numerical differentiation, dynamic ON-resistance, semiconductor

1 Introduction

Separation of exponentials, i.e. lifetime extraction from multi-exponential measurement data represented by linear combinations of decaying exponential functions with various lifetimes, is a classical area of research, related to the inverse Laplace transform and the problem of weighted moments, with numerous applications and a vast literature spreading over a century; see, e.g., [11, Chap. IV] and references there. We are interested in a particular case, where the data, the lifetimes, and the corresponding coefficients in the linear combinations are all real-valued, in contrast to a typical scenario of multiple signal classification for frequency estimation and emitter location. One additional assumption is that the lifetimes are widely spread, by orders of magnitude, practically requiring a time-logarithmic scale to represent the measurement data, which rules out traditional methods utilizing uniform time grid. Our particular application relates to analyzing trapping and detrapping of charge carriers, i.e. electrons and electronic holes, in electronic devices, as in [7,8]. Trapping an electron means that the electron is captured from the conduction band, while detrapping an electron conversely means capturing a hole from the valence band; see [10, Chap. 11].

* Accepted to ICDM 2016. An extended preliminary version posted at arXiv.

** Andrew.Knyazev@merl.com, http://www.merl.com/people/knyazev

In semiconductor devices, traps pertain to impurities or dislocations that capture the charge carriers, and keep the carriers strongly localized; see, e.g., [1,13,15,16]. Effects of the traps on performance of the semiconductor devices are temporal and eventually decay over time, i.e. the behavior of the device stabilizes and the measured quantity, e.g., the ON resistance, approaches a constant. Each trap can be assumed to behave exponentially decaying in time t with a specific lifetime τ, such as a purely exponential process in time t described by a function $c_\tau e^{-t/\tau}$, in a process reversed to a charge carrier avalanche. The parameter τ is commonly called the lifetime, and is inversely proportional to the decay rate, since $d(c_\tau e^{-t/\tau})/dt = -c_\tau e^{-t/\tau}/\tau$. The coefficient (also called the magnitude) c_τ expresses the magnitudes of the purely exponential process $c_\tau e^{-t/\tau}$, and represents the initial, at the time $t = 0$, contribution of the corresponding process to the measured data. For semiconductor traps, the magnitude c_τ may be related to the total initial, at the time $t = 0$, number of charges participating in the trap process for the fixed lifetime τ, described by the exponential process $c_\tau e^{-t/\tau}$. Depending on the sign of the charge of the carrier and the measured quantity, a single trapping process may lead to an increase with $c_\tau < 0$, called detrapping, or a decrease with $c_\tau > 0$ of the measured quantity.

Collectively, multiple traps are assumed to independently influence the operation of the device in an additive fashion with possibly various lifetimes τ and initial concentrations c_τ of the carriers in every trap. Therefore, the measurement data can be assumed to be multi-exponential in time,

$$I_{data}(t) \approx \sum_{i=1}^{n} c_i e^{-t/\tau_i} + I_\infty, \tag{1}$$

where one needs to determine the number $n > 0$ of terms in the sum, the positive lifetimes τ_i and the corresponding nonzero coefficients c_i, which can be positive in trapping and negative in detrapping processes. In our application, device recovery from the traps can take nano-seconds, minutes, or even days, therefore both t and τ need to be presented in the time-logarithmic scale, e.g., in base 10 we substitute $t = 10^s$ and $\tau = 10^\sigma$. We thus need to determine the values of σ_i and the corresponding nonzero coefficients c_i such that

$$I_{data}(s) \approx \sum_{i=1}^{n} c_i e^{-10^{s-\sigma_i}} + I_\infty, \tag{2}$$

e.g., by minimizing the least squares fit.

Since the traps may act at the nano-scale and the measurements are performed at the macro-scale, the number n of terms in the sum, if every single trap is given its own index, is practically infinite, i.e. the sum mathematically turns into the integral

$$I_{data}(t) \approx \int_{\tau_{\min}}^{\tau_{\max}} c_\tau e^{-t/\tau} d\tau + I_\infty, \tag{3}$$

where one needs to determine the interval of the lifetime values $0 < \tau_{\min} < \tau_{\max}$ and the function c_τ. Formula (3) clearly reminds us of the Laplace and the

inverse Laplace transforms, leading to some exponentially ill-posed numerical problems; see, e.g., [4]. Switching to the time-logarithmic scale, one can make a formal substitution $t = 10^s$ and $\tau = 10^\sigma$ in integral (3), or consider the integral analog of the sum appearing in (2) as follows

$$I_{data}(c) \approx \int_{\sigma_{\min}}^{\sigma_{\max}} c_\sigma e^{-10^{s-\sigma}} d\sigma + I_\infty, \tag{4}$$

where one needs to determine the function c_τ, while the interval of the lifetime logarithm values $[\sigma_{\min}, \sigma_{\max}]$ can in practice be often selected *a priori*.

The analysis of the recovery is extremely important as the traps severely degrade the performance and reliability of semiconductor devices. Trap analysis is also important for characterizing the formation and behavior of the traps so that the devices can be modeled, designed, and manufactured with improved performance and reliability. The lifetimes are affected by material temperatures and activation energies of the traps. The captured or released coefficient could be a function of the initial number of the traps to be filled or number of carriers in the traps to be released, respectively. Methods that can extract the lifetimes of the trapping and detrapping processes from the measurement data allow detecting and analyzing the traps. The measured data, e.g., the ON resistances, are undoubtedly noisy. The noise distribution function over time is unclear, except that measuring in a short time range is practically difficult, so one may expect larger measurement errors, compared to a long time range.

Having the transient data, measured as a function $I_{data}(t)$ of time t, one goal can be to determine the constant I_∞, the total number n of purely exponential components present, and numerical values of the lifetimes τ_i and the corresponding magnitudes c_i for every exponential component in (1), wherein the value n is chosen as small as possible while fitting the data within a given tolerance. In the time-logarithmic scale, we determine σ_i in (2), having the data measured for some values of s. Such a problem can be called a "discrete lifetime extraction," where the extracted lifetimes represent the dominating lifetimes, averaged over all actual physical traps at the nano-scale.

An alternative goal can be to determine the constant I_∞, the interval of the lifetime values $[\tau_{\min}, \tau_{\max}]$, and the function c_τ, describing the distribution of the lifetimes in (3). In the time-logarithmic scale, instead of c_τ we determine the function c_σ in (4) describing the distribution of the logarithms of the lifetimes, having the data measured as a function of s. Such problems can be called "continuous lifetime extraction." Computationally, continuous lifetime extraction problem (4) can be approached by introducing some grids for the s and the σ in the time-logarithmic scale, approximating the integral (4) using a quadrature rule, and solving for c_σ on the grid for the σ.

We note that the problem of multi-exponential extraction is key in many applications, e.g., fluorescence imaging and magnetic resonance tomography, but common ranges of the exponents and measurement times may not be nearly as large as in our application of analyzing traps in semiconductors that requires using the logarithmic scale, both for the exponents and the measurement times.

2 Differentiating exponential functions in the time-logarithmic scale

— I differentiate you!
— I am e^t...

What possibly exciting would one find by looking at the first derivative of exponential functions? The answer is surprising, if the time-logarithmic scale is used. In Fig. 1, several basis functions from the linear combination (1) with $\tau_i = 1$, 10, 100, 1000 and their corresponding derivatives are plotted in the time-logarithmic scale.

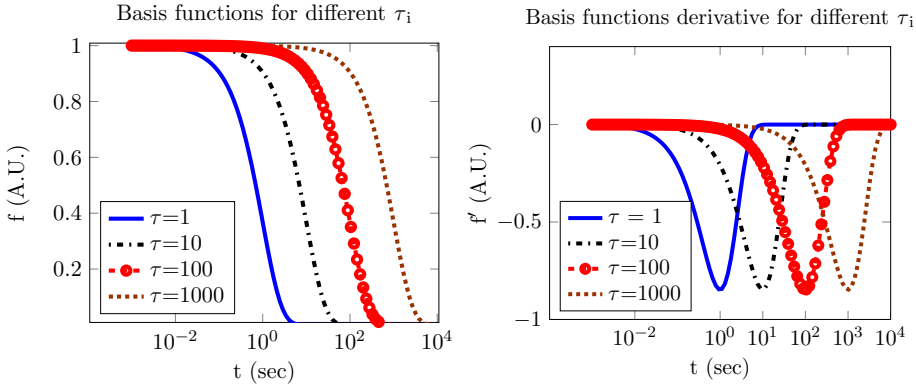

Fig. 1: The exponential basis functions and their analytic derivatives in the time-logarithmic scale

We notice in Fig. 1(b) that the peaks of the absolute values of the first derivatives of the basis function show us the locations of the lifetime constants τ! This can also be proved analytically using the following expression for the first derivative the time-logarithmic scale

$$\frac{d\left(e^{-t/\tau}\right)}{ds} = -\log(10)\, 10^{s-\sigma} e^{-t/\tau}, \text{ where } t = 10^s \text{ and } \tau = 10^\sigma \tag{5}$$

by direct calculation of one more derivative, which we leave as an exercise for the reader. Formula (5) gets simplified to $d\left(e^{-t/\tau}\right)/ds = -e^{s-\sigma} e^{-t/\tau}$, if $t = e^s$ and $\tau = e^\sigma$, but we use base 10 to follow the original format of the measurement data. Therefore, for a purely exponential, i.e. with $n = 1$, data curve (1), its only lifetime τ_1 can be easily and quickly determined by locating the point of the maximum of the first derivative.

We also observe in Fig. 1(b) that the first derivative of every basis function is reasonably well localized around the corresponding value of the lifetime

constant τ, quite quickly vanishing away from τ, especially to the right. The vanishing can be explained analytically by the behaviors of the multiplier $10^{s-\sigma}$ if $t = 10^s < \tau = 10^\sigma$ and the multiplier $e^{-t/\tau}$ if $t = 10^s > \tau = 10^\sigma$ in formula (5) of the first derivative. This crucial observation has three important implications for designing the lifetime extraction methods in the time-logarithmic scale.

First, in the multi-exponential case $n > 1$ in (1), assuming discrete lifetimes, as in Section 5, the local picks of the absolute of the first derivative of the data curve can reveal the total number n of dominant lifetimes and their approximate locations. Multi-exponential transient spectroscopy (METS) in [14,12] is based on multiple differentiation in the time-logarithmic scale to approximate the dominant lifetimes. However, already even the second order numerical derivative is found to be too sensitive to measurement noise and inexact computer arithmetic in our numerical tests and unable to provide reliable lifetime estimates, even combined with low-pass filtering. Thus, we identify the number of the dominant lifetimes and their approximate values to be used as initial approximations for the nonlinear least squares method described in Section 5 typically using only the first derivative. We note that Pade-like approximations to the lifetimes are shown in [17] to include classical Prony's method [3] and advocated in [5] as being more robust compared to METS, but use the uniform, rather than logarithmic, time scale, unsuitable for our needs.

Second, in the derivative space, the basis of the first derivatives of the functions, plotted in Fig. 1(b), compared to the basis of the functions themselves, may behave better, having a bit smaller condition number of a Gram matrix, which implicitly appears in the linear least squares fit method in the derivative space. Thus, the linear least squares fit of the first derivative of the data curve using the first derivatives of the basis function in the derivative space may expected to be more numerically stable and can be performed on a computer with the standard double precision floating point arithmetic with larger n, in contrast to the linear least squares fit of the the data curve itself using the original basis function, descried in Section 3 following [7,8,9]. The continuous lifetime extraction in the logarithmic scale using dense, e.g., uniform, grids in the logarithmic scale is somewhat more numerically feasible in the first derivative space, as we mention in Section 4. We can also take advantage of the first derivatives of the basis functions by adding a weighted quadratic term containing the first derivatives to the function minimized by the nonlinear least squares fitting.

3 Continuous lifetime extraction on a uniform lifetime grid in the time-logarithmic scale by del Alamo et al.

The methodology of [7,8,9] to extract time constants of the dominant traps is essentially based on an assumption of a continuous distribution of lifetimes, approximating the integral in (4) by a sum in (2) with a large number n using uniformly spaced grid on a given interval $[\sigma_{\min}, \sigma_{\max}]$ of the logarithms σ_i of the lifetimes τ_i with unknown coefficients c_i. The trapping and detrapping transient data, $I_{data}(t)$, is analyzed by fitting the data to a weighted sum of pure expo-

nentials in (1) using the linear least squares method. The fitting is performed to minimize the sum of $|I_{data} - I_{fitted}|^2$ at the measured points, where the magnitudes c_i are the fitting parameters to be computed, whereas the lifetimes τ_i are predefined. For the fitting functions in [7,8,9], $n = 100$ exponentials are typically used with the lifetimes τ_i that are equally spaced logarithmically in time. Positive (negative) values of c_i correspond to the trapping (detrapping) processes and each c_i represents the magnitude of a trapping (detrapping) process with respect to the lifetime τ_i. The extracted values of the magnitudes c_i can be used to construct a time-constants spectrum by plotting them as a function of τ. The vertical axis is in terms of arbitrary units (A.U.).

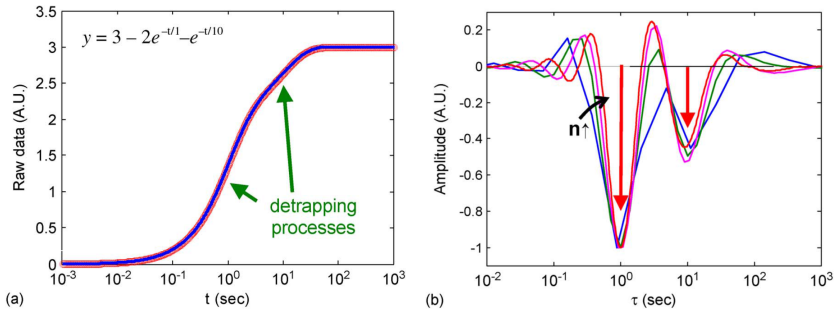

Fig. 2: Examples from [9]. (a) Time-domain signal of a synthetic current transient (red:data, blue:fitted curve). (b) Time constant spectrum extracted from the fitting of the time-domain signal with various numbers of exponentials ($n = 20, 50, 100, 500$).

In [9], the synthetic data curve $y = 3 - 2e^{-t/1} - e^{-t/10}$ is used to demonstrate the methodology as shown in Fig. 2(a). The time domain signal in this example is composed of two pure exponential detrapping components with lifetime constants $\tau_1 = 1$ and $\tau_2 = 10$. The corresponding coefficient for the lifetime constants are -2 and -1, correspondingly. Various numbers of exponentials, $n = 20, 50, 100$, and 500, are tested to perform the linear least squares fitting. The calculated amplitudes are plotted in the time-constant spectrum in [9, Fig. 2] and reproduced here in Fig. 2(b). The time-constant spectrum lifetime distribution reveals the two exponential components and their relative amplitudes. With $n = 20$, the maximum errors in the lifetime and amplitude are 11% and 5%, correspondingly. As the number n of the grid points for τ increases, the time constants and amplitudes may become more accurate, but at the cost of more computing time for the fitting process, according to [9].

Locating the actual values $\tau_1 = 1$ and $\tau_2 = 10$ by checking the peaks of the computed time-constant spectrum lifetime distribution in Fig. 2(b) requires some educated guessing and multiple calculations with various n, since a few evidently erroneous distortions clearly appear in the time-constant spectrum

in Fig. 2(b). These errors are explained in reference [9, p. 134] by the following two reasons. First, the number n of exponential components is used for the fitting may be not large enough. Second, the basis functions used in the sum (1) are not orthogonal to each other. The least squares fitting as formulated in [7,9] is a simple quadratic minimization problem. However, in order to prevent over-fitting that makes the time-constant τ spectrum numerically unstable, some constraints, such as lower and upper bounds or smoothness in the spectrum, have been added. Due to such constraints, the numerical minimization takes especially long time to be performed as the dimensionality n of the minimization problem increases. The typically chosen $n = 100$ number of terms in (1) represents a practical compromise between the computation time and the meaningfulness of the computed result, according to [9].

Computer codes used in [7,8,9] are not publicly available, so we are not able to calculate ourselves the numerical results presented in [7,8,9] and reproduced in the present paper. Our theoretical explanation of the computational difficulties of the methodology proposed and used in [7,8,9] is based on the following numerical analysis arguments. On the one hand, a good fitting requires the number n of the basis functions to be large enough. On the other hand, the exponential basis functions in (1) are not just non-orthogonal, but in reality nearly linearly dependent for large n. A numerically invalid basis thus may be formed for large n, leading to an extremely poorly conditioned least squares problem, resulting in erroneous lifetime computations due to unavoidable computer round-off errors and the measurement noise in the data. Applying the proposed constraints to numerically stabilize the least squares minimization may dramatically increase the computational time, while still cannot satisfactory resolve the inaccuracy in the lifetime calculations. In the next section, we summarize our own experience extracting the continuous lifetimes on a dense grid using the least squares, without any constraints.

4 Continuous lifetime extraction in the time-logarithmic scale revisited and regularized

The case of the time-logarithmic scale is not common in the literature, so in this section we present our heuristic experience making a naive version of the continuous lifetime extraction in the time-logarithmic scale to work using the linear least squares applied to (2) without any constraints for reasonably large values of n for several representative examples of synthetic data, using standard double precision computer arithmetic and off-the-shelf software libraries. It is important to realize that solving the integral equation of the first kind, given by (4), using its approximation (2) with a large number n on a dense grid on an interval $[\sigma_{\min}, \sigma_{\max}]$ of the logarithms σ_i of the lifetimes τ_i with unknown coefficients c_i and utilizing a dense grid of the logarithms s of the times t of data measurements, is clearly an ill-posed problem, possibly extremely sensitive to details of its setup and numerical procedures.

The first choice that needs to be made is related to an experiment design. We conjecture that an optimal grid for the measurements is a uniform grid of the logarithms s of the times t of data measurements on an interval $[s_{\min}, s_{\max}]$ that needs not only include all the values of σ_i corresponding to significantly nonzero coefficients c_i in the synthetic data, but in fact also have some more room, so that at the both end points of the interval $[s_{\min}, s_{\max}]$ the measurement data $I_{data}(s)$ behaves like a constant, having a small absolute value of the first derivative $I'_{data}(s)$ with respect to s. For example, if the synthetic data curve is generated by a Gaussian function c_σ centered at the mean point 0 with the standard deviation 1, a good smallest interval $[s_{\min}, s_{\max}]$ can be $[-4, 4]$ for the regularized (described later) linear least squares or $[-5, 5]$ for the standard linear least squares, to be reasonably numerically stable.

The second choice is the type of the grid for σ, where we advocate again using a uniform grid of the logarithms σ of the lifetimes τ on an $[\sigma_{\min}, \sigma_{\max}]$. Interestingly, the interval $[\sigma_{\min}, \sigma_{\max}]$ cannot be chosen seemingly naturally to be the same as the already decided interval $[s_{\min}, s_{\max}]$. The reason becomes clear from Fig. 1. The basis function $e^{-10^{s-\sigma}}$ in Fig. 1(a) and its derivative in Fig. 1(b) can get chopped being restricted to the interval $[s_{\min}, s_{\max}]$, no longer representing a desired shape. If the σ is chosen to be within a smaller interval, approximately $[\sigma_{\min}, \sigma_{\max}] = [s_{\min} + 2, s_{\max} - 1]$, for most of the shape of the basis function $e^{-10^{s-\sigma}}$ to fit the interval $[s_{\min}, s_{\max}]$, then we may have not enough range in σ to approximate the actual values of the logarithms σ of the lifetimes τ present in the data. Alternatively, we can choose a larger interval, approximately, e.g., $[\sigma_{\min}, \sigma_{\max}] = [s_{\min} - 1, s_{\max} + 1]$, but then we run into a different trouble of having nearly constant restrictions of the basis function $e^{-10^{s-\sigma}}$ and its first derivative. The latter trouble, however, can be detected and the enlarged interval $[\sigma_{\min}, \sigma_{\max}]$ gets the final tuning by chopping at both ends. In our tests, the tuned enlargement of the interval $[\sigma_{\min}, \sigma_{\max}]$ appears to be more numerically stable, compared to making the interval smaller.

Finally, we need to decide the grid size, which can be different for s and σ. The balanced choice, i.e. the same step lengths of the grids for s and σ, appears to be most numerically stable. Since the suggested interval $[\sigma_{\min}, \sigma_{\max}]$ is only a bit longer, if enlarged, compared to the interval $[s_{\min}, s_{\max}]$, this translates in about the same number of grid points for s and σ. For example, for $[s_{\min}, s_{\max}] = [-5, 5]$ we numerically get reasonably robust results for up to values 200.

However, if the measurements have been already made, it is the number of the measurements that can determine the number of grid points for s on the interval $[s_{\min}, s_{\max}]$. If this number is greater than the largest possible balanced value 200, the number of the grid points for σ has to be correspondingly reduced, to keep the computations numerically stable. Interpolating the measured data to a uniform grid with a smaller number of grid points may create distortions.

To that end, let us describe the actually used standard and regularized linear least squares computational procedures. Using the already decided grids S for s

and Σ for σ, correspondingly, two matrices are calculated,

$$F_0 = e^{-10^{S-\Sigma}} \text{ and } F_1 = -\log(10)10^{S-\Sigma}\, e^{-10^{S-\Sigma}}. \tag{6}$$

The numerical sanity of calculations is checked by computing the condition numbers of the matrices F_0 and F_1. The column vector C, formed by the coefficients c_i is determined by the linear least squares fit of F_0C to $I_{data}(s)$ (assuming for simplicity of presentation that $I_\infty = 0$) or by fitting F_1C to $I'_{data}(s)$.

The matrices F_0 and F_1 have a special structure of a Toeplitz (diagonal-constant) matrix. Thus, the linear least squares fit can be performed using specialized methods for Toeplitz matrices. Numerous known fast methods for Toeplitz matrices are often unsuitable for ill-conditioned systems like ours. Since the sizes of Toeplitz matrices of practical importance for our application cannot be that large, the considerations of performance are secondary for us, compared to accuracy. We have implemented an SVD-based Tikhonov regularization and smooth thresh-holding as in, e.g., [6], although it does not actually take advantage of the Toeplitz structure. The choice of the regularization parameter can be done by hand (especially well for synthetic data, where the answer is known!). The Tikhonov regularization stabilizes the computation and typically may accurately determine the general shape of the distribution c_τ, but often gives inaccurate amplitudes, smoothing down the peaks.

Overall, under the assumptions and with the suggestions described in this section, we are able to use our linear least squares computational procedures to solve reasonably challenging synthetic data problems in negligible computer time with good accuracy in the "eye-ball norm". The computational results using the derivative-based fit matrix F_1 may be more numerically stable for noiseless synthetic data curves, compared to original data based matrix F_0. The code performs especially well for smooth distributions c_τ, but can also handle purely discrete cases although resulting in some artifacts near the discontinuity points of the function c_τ. However, the accuracy of the solution in not that good on our actual practical data sets, resulting in noisy without the regularization or strongly depending on the choice of the regularization parameter lifetime distributions, without an evident protocol on choosing the regularization parameter.

One possibly most important apparent reason for such a drop in accuracy for practical measurements is that the range for the measurement time in logarithmic scale $[s_{\min}, s_{\max}]$ is not large enough for the measured quantity to become stabilized on both ends of the interval and to include the range of the present lifetimes somewhat extended, as required for numerical stability of calculations. On the right end $s_{\max}$ of the interval, if the data curve has already approached a constant and no lifetimes are anticipated with $\sigma > s_{\max}$ it is easy to extend the interval $[s_{\min}, s_{\max}]$ to the right, substituting a larger value for $s_{\max}$ and extending the data curve using the constant at $s_{\max}$. This approach works well on synthetic data, but does not improve the situation much with the practical data available to us. The main difficulty is with the data at the left end $s_{\min}$ of the interval, where the data curve apparently cannot approach a constant for physical reasons in our application. Indeed, our measurement data represent

the influence of the traps in semiconductors. The value of $s_{\min}$ is determined by limitations of existing sensor technologies, the smallest being approximately 10^{-7} in our data. There are no physical reasons to expect an absence of the traps with lifetimes $\tau < 10^{s_{\min}} \approx 10^{-7}$. However, as we have discussed above, due to numerical stability issues of the basis of the exponents, there are limitations on how small $\sigma_{\min}$ can be, given $s_{\min}$. Traps with $\log_{10} \tau$ in the interval $(s_{\min} - 2, \sigma_{\min})$, if present, affect the measurement data strong enough to make the outcome of the exact linear least squares fit wrong.

Our main contribution is changing the paradigm of the continuous lifetime extraction to the discrete one, eliminating the difficulties of the linear least squares fit, described in this section, as well as faced in [7,8,9]. Next, we present our two independent key ideas reformulating the fitting problem to make the numerical solution quick and reasonably accurate for synthetic and practical data.

5 Discrete lifetime extraction: brief description and advantages

Using non-orthogonal basis functions (1) would be numerically more stable in the least squares fitting if the lifetimes τ_i in (1) were *discrete*, i.e. if n was small and the lifetime values τ_i were sparsely distributed separated enough from each other. Both these requirements are violated in the methodology of [9], enforcing the uniform placement of the lifetimes τ_i, essentially assuming a continuous distribution of lifetimes τ in the integral in (3) approximated by a sum in (1). In contrast, for the purpose of computational stability and efficiency, let us assume the discrete distribution of the lifetimes τ_i in (1), and consider possible approaches to the computational lifetime extraction.

We note that the discrete lifetime extraction problem is a particular case of a standard parameter estimation problem, with vast literature. The classical Prony method and its numerous extensions require a constant time sampling interval leading to an equidistant grid in time t, which is impractical for our application, where the time t can range from 10^{-7} to 10^4 seconds and the measurements are given in the time-logarithmic scale. Another classical approach, the exponential peeling method, does not technically require the measurement points in time to be equidistant and can extract the lifetimes one after the other, having thus a disadvantage of an increasing and unrecoverable error after every performed extraction. Many methods, especially in audio-related applications, such as multiple signal classification for frequency estimation and emitter location, are specifically aimed at the case, where the data and all parameters in (1) are complex-valued, not taking advantage of our purely real case. Some recent approaches, popular in compressed sensing, minimize the sparsity of the lifetimes at the costs of heavy computations, allowing an optimal resolution determining closely located discrete lifetimes, which is irrelevant in our application.

We choose a generic minimization software, e.g., available in Octave and MATLAB, performing the iterative nonlinear least squares fit of the data to the sum (1), varying all parameters in (1) using real computer arithmetic. Speed up

convergence of the iterations of the minimizer, we explicitly derive the Jacobian of the minimized least squares fit function and provide it to the minimization software. Any iterative nonlinear least squares method is more computationally difficult, compared to the linear least squares, and typically requires reasonable initial approximations for all the parameters in (1) for fast and accurate computations. We discuss our choice of the initial approximations in the next section.

In Fig. 3, we demonstrate an application of our fitting algorithm to the same synthetic data curve as in Fig. 2 to illustrate our methodology. The comparison between the fitting curve and the synthetic data is shown in Fig. 3(a); the error between them is also shown to be bounded by 7×10^{-6}. The time-constant spectrum from our algorithm are shown in Fig. 3(b). Our time-constant spectrum clearly and accurately demonstrate the two exponential components (1 and 10) with their absolute values of corresponding coefficient (-2 and -1). There are no undesirable false lifetimes in our time-constant spectrum. Therefore, no extra effort is required to further process the time-constant spectrum, which might be needed in [9]. The maximum error in our time constant and corresponding coefficient fitting is 6×10^{-6} in this example.

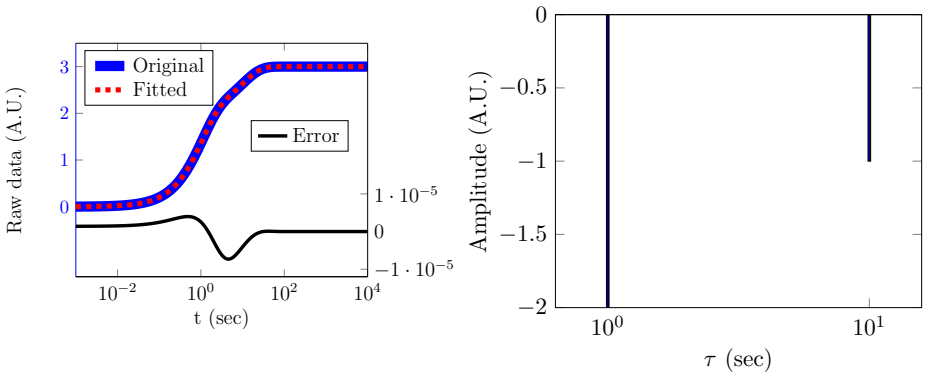

Fig. 3: Our discrete time-constant analysis: (a) Time domain signal of the same synthetic current transient (Blue:data, red:fitted curve, black:error). (b) Time constant spectrum extracted by our method.

The advantages of our methodology to extract the discrete lifetimes in this example, as seen in Fig. 3, are as follows.

- The time-constant spectrum of our method clearly and accurately demonstrates the two lifetimes and their corresponding coefficients as appears in the synthetic function, in contrast to the time-constant spectrum of [9] as shown in Fig. 2. It is difficult to determine whether the small peak as shown in Fig. 2(b) is an actual lifetime constant or an error.

- o The magnitudes corresponding to lifetimes are the actual coefficients' values whereas the magnitudes in [9] only reflect the ratio between the actual values of coefficients.
- o Our method is numerically stable and no constraint is applied during the optimization. In contrast, constraints such as lower and upper bounds or smoothness in the spectrum have been added in [9] in order to prevent over-fitting. Determining reasonable constraints may be difficult. Constrained minimization can be slow.
- o Our method is highly computationally efficient. It takes less than a second to calculate the time-constant spectrum for the synthetic function as shown in Fig. 2(a) while the code in [9] takes hours.

6 Discrete lifetime extraction: detailed description

The details of our algorithm are as follows.

1 In Fig. 4 101, interpolate the data using a dense uniform logarithmic grid in time domain with the grid size of 2^n. The measurement data are usually not measured uniformly in log scale. This step is a preparation for the following steps because discrete Fourier transform will be used to handle the measurement noise and a uniform grid is needed. Grid size of 2^n will improve the FFT from $O(N^2)$ to $O(NlogN)$ by utilizing the well-known radix-2 Cooley-Tukey algorithm [2].

2 In Fig. 4 102, filter out the noise in the measurement data. There are two alternative approaches to accomplish this in our algorithm. First one, perform a cosine Fourier transform to obtain the cosine Fourier spectrum. By analyzing the cosine Fourier spectrum and applying a proper filter in frequency spectrum, the noise in the measurement data can be filtered. Second one, apply smoothing to the data curve with measurement noise to obtain a smooth data curve.

3 In Fig. 4 103, obtain the derivative of the data curve. We implement two alternative approaches to accomplish this. First one, we transform the data into Frequency domain utilizing the Fast Fourier Transform (FFT), obtain the derivative using the Fourier spectral method, and perform inverse FFT to find the derivative in real time domain. Second one, numerically calculate the derivative using the finite difference method.

4 In Fig. 4 104, identify the nearby positions of the dominant trap lifetimes τ_i in time-constant spectrum by finding the maxima of the absolution value of derivative. Each maximum in the absolute value of derivative corresponds to an exponential component with the lifetime constant near that location. This step will give good initial approximation for the values of trap lifetimes (not only the locations, but also the number of trap lifetimes).

5 In Fig. 4 105, use linear least squares method to find the values of the coefficients c_i corresponding to the trap lifetime. This step will give a good initial approximation for the coefficients c_i.

6 In Fig. 4 106, use all the above lifetimes and coefficients as the initial approximation and employ the non-linear least squares method to optimize both c_i and τ_i simultaneously for all $i = 1, \ldots, n$. This step will improve our results in the previous few steps for τ_i and c_i.

The flowchart of our algorithm is shown in Fig. 4.

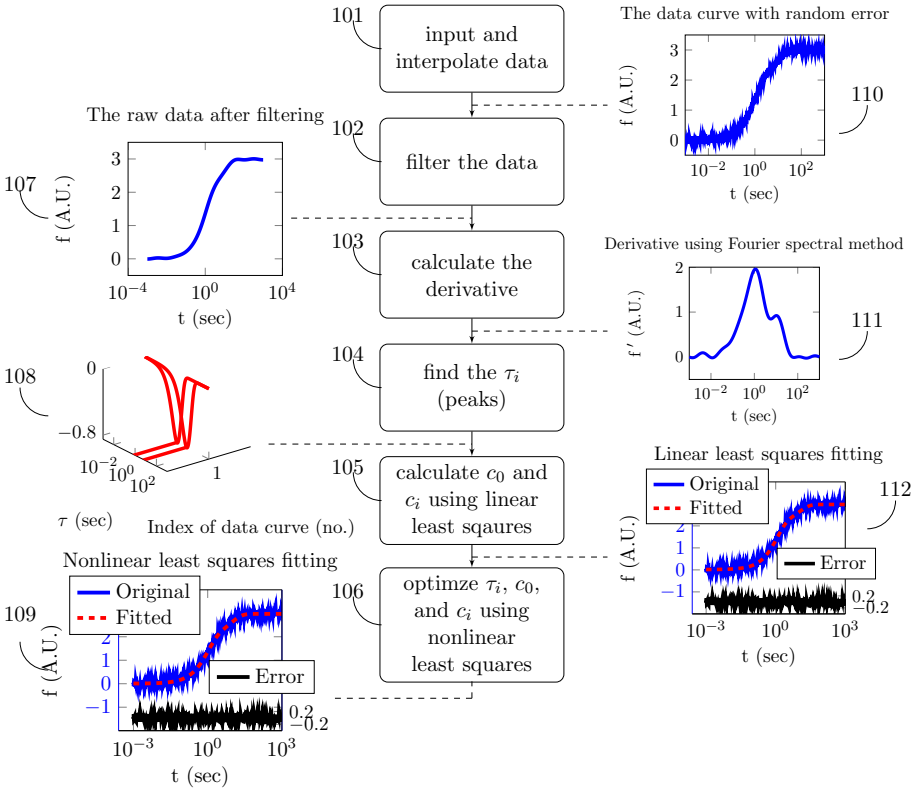

Fig. 4: The flowchart of our algorithm

We can also perform simultaneous analysis of multiple data curves, with the same lifetimes, but various coefficients, which is beneficial, e.g., if the same device is tested multiple times under slightly different conditions. The derivative of every data curve separately is used to determine potential lifetimes, that are then consolidated into a single set for all curves, using thresh-holding and averaging of nearby values. The remaining steps perform the least squares fits for all curves together, increasing robustness of calculations.

7 Application on experimental data representing dynamic ON-resistance

We finally test our algorithm on experimental data from [8, Fig. 2]. We obtain in Fig. 5 a similar data fit and the envelop shape of our discrete time-constant spectrum agrees with the continuous one obtained in [8, Fig. 2].

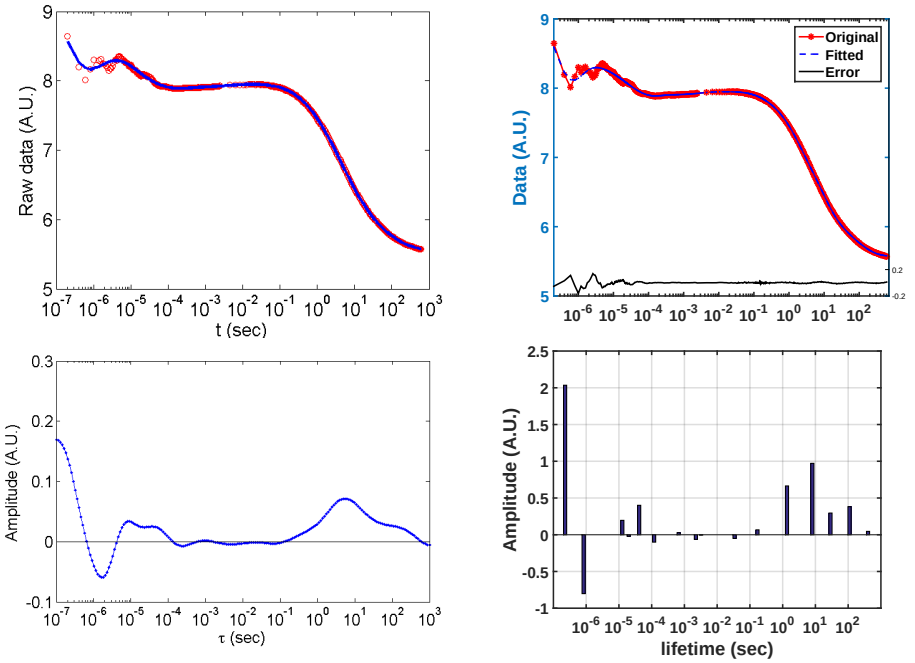

Fig. 5: The original data, the fitted curve, and error (top) and time-constant spectrum (bottom) using continuous [8] (left) vs. our discrete (right) fit.

Conclusion

We extract trap lifetimes and corresponding coefficients from experimental data, in time logarithmic scale, of dynamic ON-resistance for semiconductors. The proposed methods are numerically stable, extremely computationally efficient, and result in high quality fitting of noisy synthetic and experimental data.

Acknowledgments

We thank Donghyun Jin and Jesus del Alamo for providing us the raw data for their [8, Fig. 2] that we use in Section 7, and the corresponding results of the fit by their method, displayed in Fig. 5.

References

1. S.C. Binari, P.B. Klein, and T.E. Kazior. Trapping effects in gan and sic microwave fets. *Proceedings of the IEEE*, 90(6):1048–1058, 2002.
2. J. W. Cooley and J. W. Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, 19(90):pp. 297–301, 1965.
3. G.C.F.M.R. de Prony. *Essai expérimental et analytique sur les lois de la Dilatabilité des fluides élastiques et sur celles de la Force expansive de la vapeur de l'eau et la vapeur de l'alkool, à différentes températures.* 1795.
4. C. L. Epstein and J. Schotland. The bad truth about Laplace's transform. *SIAM Review*, 50(3):504–520, 2008.
5. R. Gutierrez-Osuna, H. T. Nagle, and S. S. Schiffman. Transient response analysis of an electronic nose using multi-exponential models. *Sensors and Actuators B: Chemical*, 61(13):170 – 182, 1999.
6. P. C. Hansen. Deconvolution and regularization with toeplitz matrices. *Numerical Algorithms*, 29(4):323–378, 2002.
7. D. Jin and J.A. del Alamo. Mechanisms responsible for dynamic ON-resistance in GaN high-voltage HEMTs. In *Power Semiconductor Devices and ICs (ISPSD), 2012 24th International Symposium on*, pages 333–336, 2012.
8. D. Jin and J.A. del Alamo. Methodology for the study of dynamic ON-resistance in high-voltage GaN field-effect transistors. *Electron Devices, IEEE Transactions on*, 60(10):3190–3196, Oct 2013.
9. J. Joh and J.A. del Alamo. A current-transient methodology for trap analysis for GaN high electron mobility transistors. *Electron Devices, IEEE Transactions on*, 58(1):132–140, 2011.
10. G. Kompa. *Basic Properties of III-V Devices — Understanding Mysterious Trapping Phenomena:*. Kassel University Press, 2014.
11. C. Lanczos. *Applied Analysis*. Dover Publications, 1988.
12. S. Marco, J. Samitier, and J. R. Morante. A novel time-domain method to analyze multicomponent exponential transients. *Measurement Science and Technology*, 6(2):135, 1995.
13. G. Meneghesso, G. Verzellesi, F. Danesin, F. Rampazzo, F. Zanon, A. Tazzoli, M. Meneghini, and E. Zanoni. Reliability of GaN high-electron-mobility transistors: State of the art and perspectives. *Device and Materials Reliability, IEEE Transactions on*, 8(2):332–343, 2008.
14. J. Samitier, J.M. Lopez-Villegas, S. Marco, L. Camara, A. Pardo, O. Ruiz, and J.R. Morante. A new method to analyse signal transients in chemical sensors. *Sensors and Actuators B: Chemical*, 18(1):308–312, 1994.
15. A. Sozza, C. Dua, E. Morvan, M.A. diForte Poisson, S. Delage, F. Rampazzo, A. Tazzoli, F. Danesin, G. Meneghesso, E. Zanoni, A. Curutchet, N. Malbert, N. Labat, B. Grimbert, and J.-C. De Jaeger. Evidence of traps creation in GaN/AlGaN/GaN HEMTs after a 3000 hour on-state and off-state hot-electron stress. In *Electron Devices Meeting, 2005. IEDM Technical Digest. IEEE International*, pages 4 pp.–593, 2005.
16. R. Stoklas, D. Gregusova, J. Novak, A. Vescan, and P. Kordos. Investigation of trapping effects in AlGaN/GaN/Si field-effect transistors by frequency dependent capacitance and conductance analysis. *Applied Physics Letters*, 93(12):124103, 2008.
17. L. Weiss and R. N. McDonough. Prony's method, Z-transforms, and Pade approximation. *SIAM Review*, 5(2):145–149, 1963.