Tolerance Bound Calculation for Compact Model Calibration Using Functional Data Analysis

SAND2019-14647C

Shahed Reza, Nevin Martin, Thomas Buchheit, Derek Tucker
Sandia National Laboratories, Albuquerque, NM 87185 USA
e-mail: sreza@sandia.gov; phone: 505-844-5327

***Abstract*—Accurate device modeling is an important prerequisite of electronic circuit design. To accurately simulate the circuit level behavior, compact model parameters for devices within a circuit must be calibrated to experimental electrical measurements. Measurements performed on a series of otherwise identical commercial off the shelf (COTS) devices reveal a distribution in their electrical behavior due to manufacturing process variation. Tolerance bounds are a useful tool for bounding a percentage of a distribution, and they can be used to characterize uncertainty in the compact model parameter extraction. The behavior of the electrical devices is typically captured as current as a function of voltage, with variability across both the current and voltage axes. This paper applies a functional data analysis approach to generate tolerance bounds on these two types of variability separately and proposes a novel approach for the estimation of transformed tolerance bounds that can be used in compact model calibration. This method is applied to simulated Zener diode data and compared with a traditional point-wise method which only considers variability along one dimension (i.e., current).**

Keywords—component, formatting, style, styling, insert (*key words****)***

I. Introduction

This Accurate device modeling is an important prerequisite of electronic circuit design. Analog circuit simulators based on SPICE (Simulation Program with Integrated Circuit Emphasis) [1], [2] used in the electrical design process support numerous canonical compact models, i.e., models for individual devices within the design. Often in high reliability and large-scale commercial designs, a series of circuit level simulations are performed to bound circuit response between well-defined tolerances to ensure performance within the required specifications.

To achieve that end, device level calibrations must also be well-bounded. Measurements performed on a series of otherwise identical commercial off the shelf (COTS) electrical devices of the same type from the same manufacturer reveal a distribution in their electrical behavior due to manufacturing process variation. Using a tolerance bound to characterize this measured distribution seems an appropriate, if not critical approach to meet the ultimate goal of simulating bounding circuit behavior. Tolerance bounds are confidence intervals on quantiles of data which can be used to determine where a certain percentage of the population fall with a given level of confidence [3]. They provide a measure of sampling uncertainty, i.e., uncertainty due to the finite sample size that was used to estimate the quantile.

There are well-established statistical methodologies for calculating tolerance bounds for scalar and multivariate data [3]. However, electrical measurement data on a single device is typically captured as current as a function of voltage. This functional data - defined as data that varies continuously across an independent variable - has two types of variability that must be accounted for - amplitude (vertical) and phase (horizontal).

This paper proposes a Functional Data Analysis (FDA) approach for establishing tolerance bounds for functional data that account for both of these types of variability. Previously we have reported the use of FDA method to select a nominal device (i.e., a device that best represents the center of the distribution) [4]. This paper expands on that work by developing methods for estimating tolerance bounds on device characterization data [5].

II. Functional Data Analysis

We have used the tolerance bounds estimation method proposed by Tucker et al. [5] and, in addition, transformed the tolerance bounds on the amplitude and phase spaces back to the original data space. The data from this additional step provides the bounding electrical behavior to be used in compact model calibration.

The first step in estimating tolerance bounds is to fit an Functional Principal Components Analysis (fPCA) model that will be used to generate functions that represent the amplitude and phase variability in the original data. fPCA is a statistical method for identifying the modes of variation in functional data [6], [7]. This fPCA model requires a characterization of the amplitude and phase variability of the data; a brief overview of these two types of variability is given below.

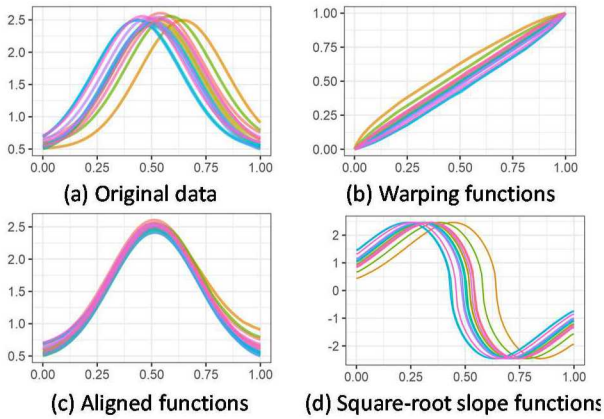

Fig. 1. Example of a Notional example of warping functional data.

A. Amplitude and phase variability

Figure 1a shows an example of functional data that contains both amplitude and phase variability. The FDA approach described in [7] provides metrics that can be used to define both of these types of variability.

First, for amplitude variability, consider F, a set of continuous, real-valued functions f on the domain $[0,1]$. Let Γ denote a set of functions γ (known as diffeomorphisms) that map from $[0,1]$ to $[0,1]$ and have the constraints of $\gamma(0) = 0$ and $\gamma(1) = 1$. The set Γ are called the warping functions shown in Fig. 1(b), that are used to align the functional data along the x axis. Aligned functions, f^*, shown in Fig. 1(c), are produced from the composition of f with γ, that is $f(\gamma)$, denoted $f \circ \gamma$.

Note that while F is defined on [0,1], this domain can be generalized to any interval on the real line.

A metric that gives the amplitude distance for two functions f_1 and f_2 is defined as

$$d_a(f_1, f_2) = \inf_{\gamma \in \Gamma} \left\| q_1 - (q_2 \boxtimes \gamma)\sqrt{\dot{\gamma}} \right\| \quad (1)$$

Here, $q(t)$ is known as the square-root slope function (SRSF), as $q(t) = \text{sgn}(\dot{f}(t))\sqrt{|\dot{f}(t)|}$, where $\dot{f}$ is the time derivative of f. $f(t)$ is transformed to $q(t)$ because the SRSFs are able to form a distance metric that is symmetric, positive definite, and satisfies the triangle inequality. More information on SRSFs are provide in [8] and [9], while details on solving Equation 1 and estimating the warping functions γ are given in [7].

For phase variability, a second metric is defined as

$$d_p(\gamma_1, \gamma_2) = d_\psi(\psi_1, \psi_2) = \cos^{-1}\left(\int_0^1 \psi_1(t)\psi_2(t)dt\right) \quad (2)$$

where $\psi = \sqrt{\dot{\gamma}}$, and $\psi \in \Psi$. This transformation on the warping functions is performed to simplify the geometry of Γ (an infinite dimensional nonlinear manifold) to a unit Hilbert sphere. To further simplify the geometry to aid in the computation of statistics, Ψ is then mapped to a tangent space:

$$T_\psi(\Psi) = \{v \in \mathbb{L}^2 | \int_0^1 v(t)\psi(t)dt = 0\} \quad (3)$$

This mapping results in a shooting vector (tangent space vector), v, that can be used to characterize the phase variability. For more details, please see [7] and [5].

These representations of amplitude and phase variability respective SRSFs (q) and the tangent space vectors (v) - will be used in the fPCA model discussed in the next section.

B. fPCA Model

To estimate tolerance bounds, an fPCA model is fit and used to sample new functions that represent the amplitude and phase variabilities. As suggested in [5], the model starts with a joint, piecewise representation of the amplitude and phase functions:

$$g^C(t) = \begin{cases} q^*(t) & t \in [0,1) \\ Cv(t-1) & t \in [1,2] \end{cases} \quad (4)$$

In Equation 4, q^* are the SRSFs of the aligned functions f^*, which isolate the amplitude variability, v are the tangent space vectors, which isolate the phase variability, and C is a scaling factor [10]. Given a sample of n functions $g_1^C, \dots, g_n^C$ with sample means $\hat{\mu}_g^C = [\hat{\mu}_{q^*}, \hat{\mu}_\gamma^C]$, the covariance matrix can be defined as:

$$K_g^C = \frac{1}{n-1}\sum_{i=1}^{n}(g_i^C - \hat{\mu}_g^C)(g_i^C - \hat{\mu}_g^C)^\intercal \in \mathbb{R}^{(2T)X(2T)} \quad (5)$$

where T is the number of points in q^* and v. Singular Value Decomposition can then be used on this covariance matrix to estimate the principal components $c_j = \langle g_i^C, U_{g,j}^C \rangle$, where i is the function and j is the principal direction. Now let $c = (c_1,\dots,c_k)$ be the k largest principal coefficients, where k is a user-specified parameter. By allowing c to vary according to a probability model, the fPCA model can be used to generate a distribution of functions. In this case, c is chosen to be distributed as a multivariate Normal distribution, with 0 mean and $k \times k$ covariance matrix Σ, defined as:

$$g^C(t) = \begin{cases} q^*(t) & t \in [0,1) \\ Cv(t-1) & t \in [1,2] \end{cases} \quad (4)$$

where $\hat{\sigma}_i^c$ is estimated from the ith eigenvalue of the sample covariance matrix defined in Equation 5.

After the fPCA model has been created, a new set of n sample functions can be generated, g_i^c, where $i = 1,..,n$. The SRSFs q_i^* and vectors v_i are extracted from g_i^c, and the vectors v_i are then transformed to the warping functions γ_i through the exponential map:

$$\gamma_i(t) = \int_0^t (\exp_{\hat{\mu}_\Psi}(v_i(s)))^2 ds \quad (7)$$

This results in a generated set of n SRSFs (representing amplitude variability) and n warping functions (representing phase variability).

C. Functional Tolerence Bounds

Recall that a tolerance bound is used to bound a percentage of the population with a given level of confidence. In this case, it is of interest to generate an equal-tailed tolerance bound, defined as the $(\alpha/2) * 100\%$ lower bound on the $(p/2)th$ quantile and the $(1-\alpha/2) * 100\%$ upper bound on the $(1-p/2)th$ quantile. Here, $(1-p) * 100$ represents the percentage of the population to be bounded, while $(1-\alpha)*100$ corresponds to the confidence level.

1) Bootstrapping to Estimate FDA Tolerance Bounds

Once the set of n SRSFs and warping functions have been sampled from the aforementioned fPCA model, the $p/2$ and $(1-p/2)$ quantiles can be estimated, denoted as $(q_{p/2}^*, q_{1-p/2}^*)$ and $(\gamma_{p/2}, \gamma_{1-p/2})$, respectively. While there are numerous ways to estimate these quantiles, the method used in this work is small extension of the method presented by Xie et al [11].

This process of generating n SRSF and warping functions and estimating the quantiles is repeated S times to get a distribution of quantile values for q^* and γ. Tolerance bounds on these quantiles can then be estimated by calculating the $\alpha/2$ and $(1-\alpha/2)$ quantiles from these distributions, respectively. This results in four tolerance bounds - $(q_{\alpha/2}^*, q_{1-\alpha/2}^*)$, the respective lower and upper tolerance bounds for q^*, and $(\gamma_{\alpha/2}, \gamma_{1-\alpha/2})$, the respective lower and upper tolerance bounds for γ. The tolerance bounds for q^* are transformed to the amplitude space via integration to produce bounds on the amplitude: $(f_{\alpha/2}^*, f_{1-\alpha/2}^*)$. Figures 2a and 2b provide examples of the respective tolerance bounds on the phase and amplitude spaces.

While the bounds on the warping functions are likely as expected, the bounds on the aligned functions may be unintuitive. This is because the amplitude bounds are constructed to capture the "shape" of the functions [5], rather than provide a point wise bound as is typical for scalar tolerance bounds. More discussion of this idea and

recommendations for other options to visualize these bounds is provided in [5], [11].

2) Transformation of Bounds

Since the goal of this work is to estimate bounds in the original data space, the bounds on the amplitude and phase spaces must be transformed. 4 pairwise combinations of the bounds are composed, that is,

- $(f_{\alpha/*2}) \boxtimes (\gamma_{\alpha/2})$
- $(f_{\alpha/^*2}) \boxtimes (\gamma_{1-\alpha/2})$
- $(f_{1*-\alpha/2}) \boxtimes (\gamma_{\alpha/2})$
- $(f_{1^*-\alpha/2}) \boxtimes (\gamma_{1-\alpha/2})$

Figure 3 gives an example of these four bounds. Again, these bounds may not be intuitive upon first glance. However, it can be seen that they bound both the phase and amplitude variability well. Two of the bounds bound the functions on the left, while the other two bound them on the right. Similarly, two bounds bound the functions from above, while the other two bound from below. All four bounds represent some combination of phase and amplitude variability that is important to capture.

III. Application on Electrical Device Data

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file.

The electrical data, collected in the form of current voltage sweeps (I-V) curves, used for this analysis was taken from 44 measurements on MMSZ5239BT1-G Zener Diodes. Further details of these measurements are given in [4] as this data represents a subset of the data used to perform FDA to find a nominal device from a set of 116 measurements. Mirroring the approach of that analysis, measurement data is converted to

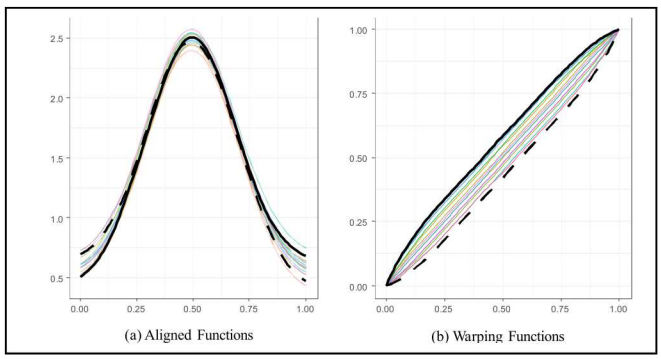

Fig. 2. Upper (solid) and lower (dashed) tolerance bounds on aligned (a) and warping (b) functions Example of a Notional example of warping functional data.

a Ln(Abs(current)) vs. Voltage scale prior to further analyses. However, using FDA to provide tolerance bounds estimates is not amenable to data with discontinuities and significant electrical noise from which to derive the SRSFs. Thus, the measured data, which had electrical noise in the very low current portions of the measurement, and discontinuities and locations where different measurement sweeps were stitched together, was replicated using a circuit simulation with the Xyce simulation tool. Simulations replicated the measured device behavior quite accurately, as shown in the example Figure 4, which gives one of the simulated diode measurements used in the FDA analysis. The simulated I-V curves also provided smooth continuous data in the low current region dominated by instrument noise, where the device response is beyond instrument detection range. If FDA is the primary goal on a set of electrical measurements, this approach could be modified to correct, or at least mitigate, the noise and discontinuities often present in most typical electrical data measurements of this type.

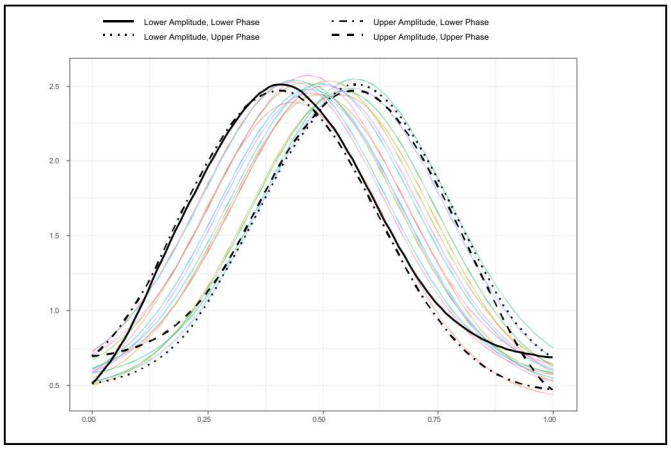

Fig. 3. Example of transformed tolerance bounds in the original data space

In the example shown in Figure 4, the log-plot of the simulation data always produces a cusp when the current transitions from positive to negative values. This cusp, an artifact of the low current simulated values and the log operation, complicates the spline fitting process. To eliminate this cusp, data below 5.1×10^{-12} A is removed from the simulated measurement and replaced with a polynomial fit, resulting in an adjustment to the data as shown by the blue dashed line in the figure. Figure 5 shows the I-V curve from all 44 simulated Zener diode devices.

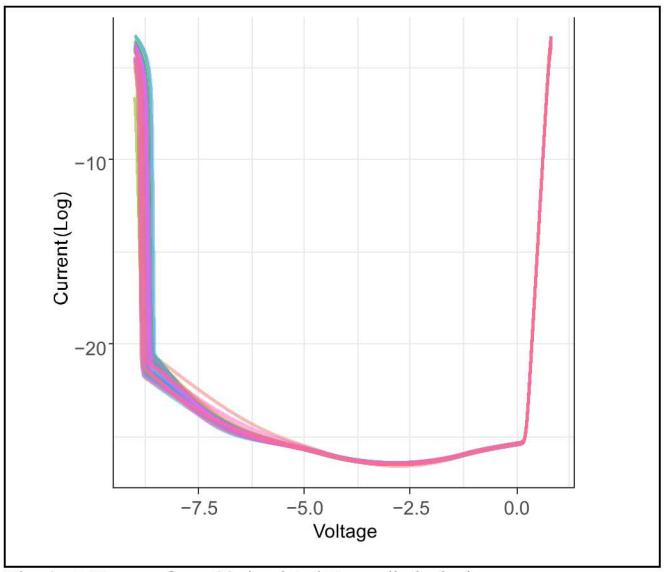

Fig. 4. I-V curve from 44 simulated Zener diode devices

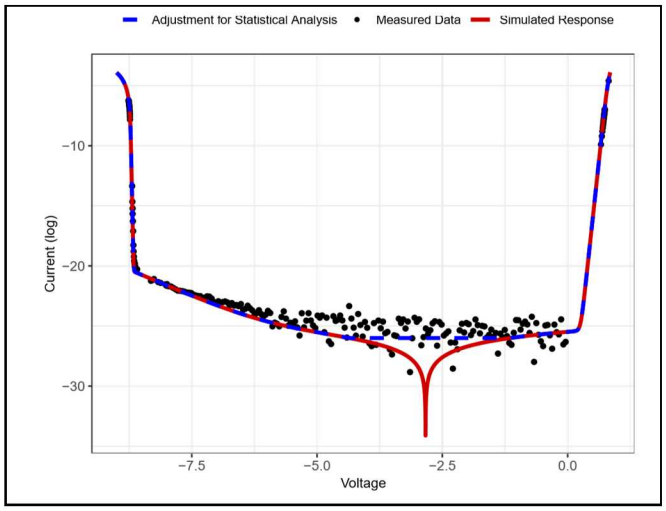

Fig. 5. Example of simulating a measured I-V curve for a MMSZ5239BT1-G Zener diode. Simulated data was used to test and demonstrate the tolerance bound analysis.

IV. Results

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file.

An fPCA model was fit to the Zener diode data using 5 dominant principal components. From the fPCA model, 500 bootstrap samples were generated to calculate tolerance bounds on the amplitude and phase spaces, as shown in Figure 6. These bounds were constructed with $p = .10$ and $\alpha = .05$ such that they represent where we would expect 90% of the devices to fall with 95% confidence. From these plots, we can see that most of the phase and amplitude variation occurs in the reverse region of the I-V curve.

A coverage study was conducted to assess whether these bounds are performing as intended. This study involved first sampling a new set of 44 diode devices from the original data with replacement (i.e., the same device can be sampled more than once). The warping functions (γ) and aligned SRSFs (q^*) were then calculated for each device. For amplitude, the 90th quantile of the SRSFs was calculated and compared to the $(\gamma_{\alpha/2}, \gamma_{1-\alpha/2})$ tolerance bound to see if the entire SRSF fell within that bound. This was repeated 500 times to estimate the confidence level of this bound (e.g., for a 95% confidence level, we would expect the SRSFs to fall within the tolerance bounds 95% of the time). This same process was performed for the phase tolerance bounds using the warping functions γ. The amplitude and phase confidence levels were estimated to be 97.7% and 99.3%, respectively. While somewhat conservative, these values are relatively close to the expected value of 95% confidence, and it is consistent with the results seen in [5].

All four pairwise combinations of these bounds were then composed to estimate the tolerance bounds in the original data space, as described in Section III-B. These bounds are shown in Figure 7. As a comparison, a point-wise non-parametric bootstrap was used to calculate a tolerance bound across

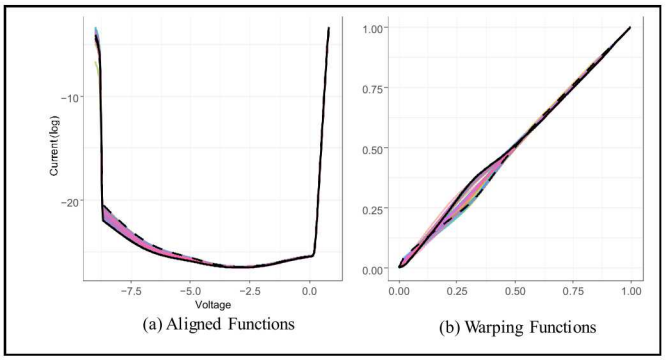

Fig. 6. Tolerance bounds on the aligned and warping functions of the Zener diode data. These functions separate the amplitude and phase variability, respectively, of the data

voltage. Ostensibly, from a use perspective, the most important region to accurately capture current and amplitude variation in the Zener diode is in the reverse breakdown region, plotted in Figures 8a and 8b. Figure 8a shows that the non-parametric point wise calculation provides bounds that, initially, seem reasonable for the given dataset of 44 measurements. However, because the point wise bound does not account for a functional relationship between the dependent and independent variables, bounds determined in this way provide a non-physical representation of real Zener diode behavior. The inflection on the bound shown on the left hand side of the data in Figure 8a is a clear example of an I-V characteristic not exhibited by any of the 44 measurements which form the basis of the analysis. Further, this characteristic is not a true physical behavior of

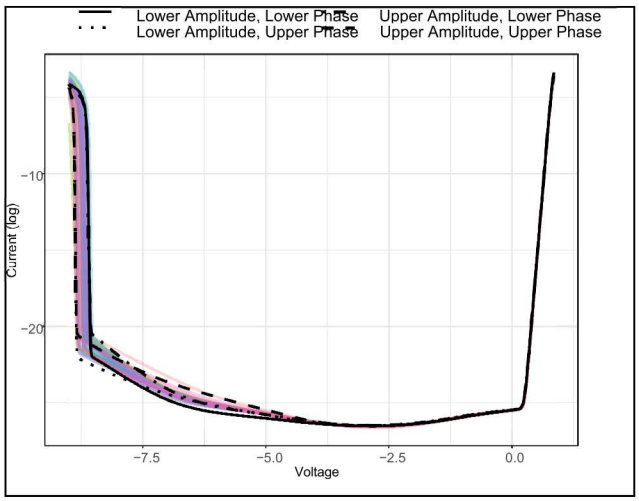

Fig. 7. Tolerance Bounds

a real diode. Extracting a set of model parameters from this behavior would lead to a non-physical representation of a real device.

Since FDA captures variability across a function and decomposes that functional variability into phase and amplitude parts, a case can be made to consider the four tolerance bounds defined by both phase variability and amplitude variability. Figure 8b provides an illustrative example. The four bounds in this region capture four bounding aspects of the diode behavior as it approaches breakdown, the knee in the curves and voltages between -8.5V and -8.85V: i) a high leakage current and high breakdown voltage,

corresponding to upper amplitude and upper phase, ii) a low leakage current and low breakdown voltage, corresponding to the lower amplitudelower phase bound, iii) a low leakage current and high breakdown voltage, corresponding to the lower amplitude-upper phase bound, and iv) a high leakage current low breakdown voltage, corresponding to the upper amplitude-lower phase bound. Each of these four bounding behaviors is important when characterizing the variability of a distribution of devices and the impact it may have in an electrical circuit.

V. VI. CONCLUSION

We presented a functional data analysis (FDA) approach to estimating tolerance bounds on a set of Zener diode devices. Tolerance bounds were first generated on the phase and amplitude spaces separately, and a coverage study was conducted to assess the performance of these bounds. The phase and amplitude bounds were then transformed to the original I-V data space, allowing for these bounds to be used in the first phase of parameter calibration for compact models in electronic circuit design applications. Applying this method to simulated Zener diode data showed that the FDA approach was able to produce several physical bounds that capture both the phase and amplitude variability in the data, particularly

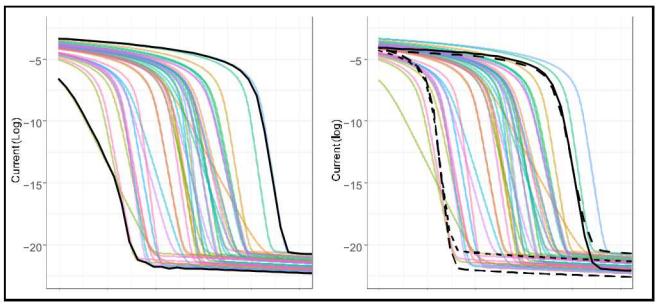

Fig. 8. Zoomed-in tolerance bounds on the reverse breakdown region for the point wise bootstrap method (a) and the FDA method (b).

in the critical reverse breakdown region. Further research is warranted to understand how bounds on the I-V space map back to the parameter space for compact model parameter calibration.

ACKNOWLEDGMENT

The authors would like to thank Biliana Paskaleva and Andrew Sandoval for providing the experimental dataset that was used in this analysis.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DENA0003525.

This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

REFERENCES

[1] A. Laha and D. Smart, "A zener diode model with application to spice2," IEEE Journal of Solid-State Circuits, vol. Vol. 16, pp. 21–22, 1981.

[2] L. Piotrowski, "An improved spice2 zener diode model for soft-region simulation capability," IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. Vol. 7, pp. 1301–1303, 1988.

[3] K. Krishnamoorthy, "Statistical tolerance regions: Theory, applications, and computation," 2009.

[4] S. Martin, T. Buchheit, and S. Reza, "Selection of a nominal device using functional data analysis," IEEE 5th Internal Conference on Data Science and Advanced Analytics, pp. 381 – 386, 2018.

[5] J. D. Tucker, J. R. Lewis, C. King, and S. Kurtek, "A geometric approach for computing tolerance bounds for elastic functional data," Under Submission, 2019.

[6] J. Ramsay and B. Silverman, "Functional data analysis," Springer, 2005.

[7] J. Tucker, W. Wu, and A. Srivastava, "Generative models for functional data using phase and amplitude separation," Computational Statistics and Data Analysis, vol. 61, pp. 50–66, 2013.

[8] A. Srivastava, W. Wu, S. Kurtek, E. Klassen, and J. Marron, "Registration of functional data using fisher-rao metric, arxiv:1103.3817v2 [math.st]," 2011.

[9] S. Kurtek, A. Srivastava, and W. Wu, "Signal estimation under random time-warpings and nonlinear signal alignment," Proceedings of Advances in Neural Information Processing Systems (NIPS), 2011.

[10] S. Lee and S. Jung, "Combined analysis of amplitude and phase variations in functional data," pp. 1 – 21, 2017.

[11] W. Xie, S. Kurtek, K. Bharath, and Y. Sun, "A geometric approach to visualization of variability in functional data," vol. 112, pp. 979–993, 2017.