IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.0429000

Compact Model Parameter Extraction via Derivative-Free Optimization

RAFAEL PEREZ MARTINEZ[1], (Student Member, IEEE), MASAYA IWAMOTO[2], (Member, IEEE), KELLY WOO[1], (Student Member, IEEE), ZHENGLIANG BIAN[1], (Student Member, IEEE), ROBERTO TINTI[3], STEPHEN BOYD[1], (Life Fellow, IEEE), and SRABANTI CHOWDHURY[1], (Fellow, IEEE)

[1]Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA
[2]Keysight Technologies Inc., Santa Rosa, CA 95403 USA
[3]Keysight Technologies Inc., Calabasas, CA 91302 USA

Corresponding author: Rafael Perez Martinez (e-mail: rafapm@stanford.edu).

This work was supported in part by the Stanford Graduate Fellowship (SGF). The fabrication tasks were performed at the nano@Stanford labs, part of the National Nanotechnology Coordinated Infrastructure, supported by the National Science Foundation under Award No. ECCS-2026822. The authors would like to acknowledge Keysight Technologies management for supporting this work.

ABSTRACT In this paper, we address the problem of compact model parameter extraction to simultaneously extract tens of parameters via derivative-free optimization. Traditionally, parameter extraction is performed manually by dividing the complete set of parameters into smaller subsets, each targeting different operational regions of the device, a process that can take several days or even weeks. Our approach streamlines this process by employing derivative-free optimization to identify a good parameter set that best fits the compact model without performing an exhaustive number of simulations. We further enhance the optimization process to address critical issues in device modeling by carefully choosing a loss function that evaluates model performance consistently across varying magnitudes by focusing on relative errors (as opposed to absolute errors), prioritizing accuracy in key operational regions of the device above a certain threshold, and reducing sensitivity to outliers. Furthermore, we utilize the concept of train-test split to assess the model fit and avoid overfitting. This is done by fitting 80% of the data and testing the model efficacy with the remaining 20%. We demonstrate the effectiveness of our methodology by successfully modeling two semiconductor devices: a diamond Schottky diode and a GaN-on-SiC HEMT, with the latter involving the ASM-HEMT DC model, which requires simultaneously extracting 35 model parameters to fit the model to the measured data. These examples demonstrate the effectiveness of our approach and showcase the practical benefits of derivative-free optimization in device modeling.

INDEX TERMS ASM-HEMT, Compact Model, Derivative-Free Optimization, Device Modeling, Diamond Schottky Diode, GaN HEMTs, Parameter Extraction, SPICE Diode Model

I. INTRODUCTION

SEMICONDUCTOR device compact models play a crucial role in the design and development of integrated circuits and systems, serving as the bridge between physical semiconductor devices and electronic design automation (EDA) tools. These models represent mathematically the electrical (and, in some instances, thermal) behavior of semiconductor devices (i.e., charges and currents) such as transistors and diodes as a function of electrical bias. Before employing these models to design circuits and systems, it is essential to diligently extract the relevant model parameters tailored to the chosen semiconductor process, ensuring that the compact model can accurately reproduce the characteristics of a specific semiconductor device. This process involves adjusting the parameters of the compact model to align with data from the semiconductor device, whether obtained through experimental measurements or simulated using Technology Computer Aided Design (TCAD) tools. However, parameter extraction has become increasingly complicated as modern compact models contain hundreds of model parameters, which are required to model the non-idealities of emerging Silicon (Si) devices such as FinFETs or III-V devices such as Gallium Nitride (GaN) high-electron-mobility transistors (HEMTs). For example, the latest version of the Berke-

ley Short-Channel IGFET Model (BSIM)-Common Multi-Gate (CMG) features over a thousand model parameters [1], whereas the most recent version of the Advanced SPICE Model for GaN HEMTs (ASM-HEMT) includes more than 200 model parameters [2].

Traditionally, manual fitting has been the default approach for parameter extraction in compact models. Given the extensive number of model parameters in modern compact models, a common strategy is to divide the complete parameter set into several smaller subsets [3]. These subsets usually correspond to specific physical elements of the device, such as drain/gate currents or junction capacitances. However, this approach typically involves a series of iterative steps and can extend over several days or weeks just to extract a single model card. It also frequently leaves engineers uncertain whether the resulting model card represents a near-optimal set of parameters or if there is room for further improvement.

Another common approach to reducing manual adjustment efforts is using gradient-based optimization methods. This involves employing a numerical nonlinear optimizer in conjunction with the Levenberg–Marquardt algorithm [4], [5] to extract model parameters [6]. However, obtaining gradient information in this context is exceptionally challenging. When conducting actual experiments, it is simply not possible to get gradients. Additionally, when simulations involve TCAD or SPICE, this task becomes very complicated. Gradient-based approaches are typically computationally inefficient, as calculating gradients (i.e., determining how a small change in each parameter affects the output) usually requires prohibitive numerical approximations.

Since extracting model parameters in semiconductor devices is a complex and time-consuming task, several approaches have been proposed due to the lack of a universal method applicable across different semiconductor technologies. Among the proposed solutions, a notable deep learning approach involves training a neural network to output the desired model parameters using the device's characteristics as inputs [7]–[12]. The device characteristics (or inputs) may include data that is either measured or simulated using TCAD, encompassing $I - V$ characteristics, S-parameters, or large-signal load-pull data. While this approach is generally effective, it demands substantial computational resources, requiring thousands of simulations to train the neural network just to extract the parameters of one type of device. This becomes particularly inefficient when fitting a new type of semiconductor device each time. The complexity of this approach is further increased by the necessity of prior parameter extraction experience to define the variation ranges of the model parameters, which is essential for generating training data for the desired device.

We propose adopting derivative-free optimization (DFO) to address the issues above. DFO methods approximately minimize a function only using the objective value (i.e., no gradients are required). They are also straightforward to implement and more computationally efficient. The primary advantage of our approach lies in its ability to identify a set of model parameters that achieves a near-optimal fit with significantly fewer simulations than would be required for a full-grid search. This approach not only mitigates the curse of dimensionality often encountered when considering tens of parameters but also allows us to explore a relatively wide range of plausible parameter values.

To further enhance the optimization process, we carefully select a loss function that addresses three key issues: 1) ensuring model performance consistently across varying magnitudes by focusing on relative errors, as opposed to absolute errors; 2) guiding the optimization process to prioritize regions of particular interest while deprioritizing less critical regions of operation; and 3) reducing sensitivity to outliers and measurement errors. Moreover, we utilize a standard model assessment method (train/test split) used in Statistics and machine learning (ML) to judge the fit of our extracted model. This method is unlike traditional approaches in device modeling that fit the model to the entire dataset, potentially leading to overfitting.

The remaining paper is structured as follows: Section II formulates the problem we are trying to solve (i.e., model parameter extraction) as an optimization problem in a more general form and introduces the DFO framework we used in the present work. Section III describes our proposed approach and the loss function we chose to tackle this problem. It also includes a straightforward example of fitting a simple two-terminal device, specifically a diamond Schottky diode, to clearly outline and effectively demonstrate the issues we are addressing. Section IV presents the modeling of a 150-nm gate length (L_G) GaN-on-SiC HEMT using the ASM-HEMT DC model, a task that involves extracting more than 30 model parameters simultaneously. Lastly, Section V concludes this article.

II. DFO FOR MODEL PARAMETER EXTRACTION

In this work, we focus on extracting semiconductor compact model parameters. This process involves identifying a set of parameter values that precisely replicate a device's characteristics, which may derive from various experiments on a fabricated device or from TCAD simulations. These characteristics can be static (e.g., $I - V$ characteristics or S-parameters), dynamic (e.g., dynamic load-lines), or a combination of both (i.e., heterogeneous).

In the context of the present work, each experiment involves collecting one or more measurements from the device. Considering k measurements, each y_i corresponds to a distinct measurement obtained from an experiment. For example, in the case of a diode, a single experiment might involve measuring multiple points on the I-V curve, where each measurement captures the current I_i for a given voltage V_i. Collectively, these measurements represent the I-V characteristics of the diode. The output of the compact model is denoted by $\hat{y}_i$ and parameterized by a p-vector $\theta = (\theta_1, \ldots, \theta_p)$, where θ is within a subset $\Theta \in \mathbb{R}^p$ that represents the selected model parameters from a feasible set. Each model parameter in θ has a range of plausible values. In some cases, it is more

convenient to work with the logarithm of the parameter's range (e.g., the saturation current in a diode). We seek to solve the optimization problem:

$$\begin{aligned} &\text{minimize} \quad \frac{1}{k}\sum_{i=1}^{k}\mathcal{L}(\hat{y}_i, y_i) \\ &\text{subject to} \quad \theta \in \Theta, \end{aligned} \tag{1}$$

where $\mathcal{L}(\hat{y}, y)$ is the loss function we want to minimize over the model parameters θ. Here, θ represents the variable in the optimization problem, $\hat{y}$ is the predicted or simulated value obtained from the compact model, and y is the true value that has been experimentally measured or simulated using TCAD.

The complexity of solving this optimization problem is that the model's output is not given by simple expressions but rather by running a SPICE simulation. This implies that obtaining gradient information is generally challenging and prohibitively expensive. This difficulty is compounded when considering tens of model parameters, which turns this task into a computationally intensive endeavor. This is a classic case of the curse of dimensionality, where exploring every potential combination of parameter values becomes impractically costly in terms of time and computational resources. In such scenarios, DFO excels at finding near-optimal solutions as this approach does not require gradients and directs the optimization effort towards promising zones of the parameter space that are more likely to yield the most accurate model fits. This significantly reduces the computational overhead by eliminating the need for calculating gradients and decreasing the number of simulations required to obtain a good fit [13].

DFO methods have also shown promise in ML, particularly in hyperparameter tuning. In the context of ML, hyperparameters are settings or configurations that control the behavior of a machine learning algorithm, such as the learning rate or the number of hidden layers in a neural network. DFO methods are effective at identifying a good set of hyperparameters that significantly enhance ML model performance [14]–[17].

Two drawbacks of DFO methods are their reduced effectiveness when dealing with hundreds of parameters and their inability to guarantee the attainment of a global solution [18]. This is primarily due to the curse of dimensionality, which makes it challenging to sample points close to the global optimum unless a large number of samples are taken. Additionally, DFO methods employ stochastic or heuristic sampling methods to explore the design space, which could lead to settling on local optima without guaranteeing a global solution. Nevertheless, the problem we are addressing is in the order of tens of parameters and results in a nearly optimal fit with far fewer simulations than what would be required by a full-grid search. This represents a clear advantage over the deep learning approaches used in previous works by [7]–[12], which require thousands of simulations to generate their training and testing datasets.

It is also worth noting that the problem we are trying to solve is analogous to calibrating a TCAD model, which involves matching experimental data with the simulated results from a TCAD device simulator by adjusting TCAD model parameters. Given that TCAD simulations are generally more expensive and time-consuming, our proposed approach is particularly advantageous. In such scenarios, DFO methods excel at finding a nearly optimal fit with far fewer simulations than a full-grid search, which is a clear advantage over manually tuning TCAD model parameters, which can take weeks or, in some instances, months to achieve a good fit.

A. DERIVATIVE-FREE OPTIMIZATION FRAMEWORK

We resort to open-source hyperparameter optimization frameworks in the present work since they employ DFO methods. Some of them include Autotune [14], HOLA [15], Hyperopt [16], and Optuna [17], to name a few. A primary advantage of these frameworks is their streamlined application programming interface (API), which enables users to configure a parameter search space with minimal coding effort. This allows for quick adjustments to the loss function or to constraints related to model parameters. The choice of optimization framework is entirely up to the user, as each framework is relatively similar and shares common features. In this work, we employed Optuna and its default sampler, the Tree-structured Parzen Estimator (TPE), for model fitting due to its capability to efficiently explore large parameter spaces. Under these settings, Optuna has demonstrated effectiveness in prior studies, navigating spaces that encompass up to 34 parameters [17], which aligns with the complexity of our task.

Most DFO frameworks follow a structured approach to improve model performance through parameter tuning. Initially, they use a sampling method to explore the parameter space and identify promising regions for optimal settings. As the process advances, these frameworks adopt refined sampling strategies, focusing on the top 20–30% of parameters that have shown the best performance. This strategic refinement enables the algorithm to gradually learn and adapt to the distribution of the most effective parameter values. As more data points are accumulated, the accuracy of identifying a superior set of parameters increases, resulting in enhanced model performance. In particular, Optuna allows the user to choose among various sampling strategies, with the default option being the TPE sampler. This sampling approach uses past outcomes to predict which parameter settings might lead to better outcomes by employing two Gaussian Mixture Models: one looks at a set of parameter values with the best results, and the other looks at the remaining parameters. The TPE algorithm then decides whether to try new and untested parameter values or use the ones already shown to work well. By balancing the search for new parameters with known effective values, this strategy enhances the likelihood of finding the best set of parameters [17]. Interested readers may refer to [19] for a comprehensive tutorial on TPE.

III. PROPOSED APPROACH FOR PARAMETER EXTRACTION

In our proposed approach, we begin by setting the device dimensions, if necessary. Following this initial step, we define

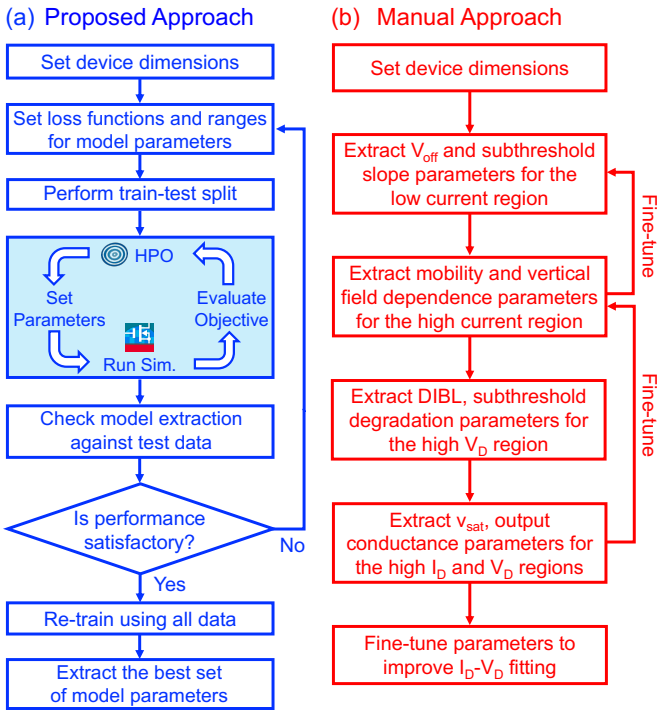


FIGURE 1. **(a) Proposed approach to extract DC model parameters using derivative-free optimization framework, and (b) Manual approach for extracting DC model parameters in the ASM-HEMT model (after [20]).**

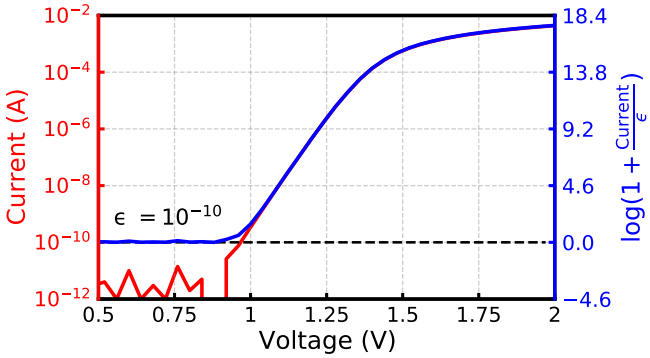


FIGURE 2. **Transformation of the loss function that effectively excludes any current values below a threshold ϵ from the fitting process. The left y-axis (red) represents the current values without the transformation, whereas the right y-axis (blue) represents the current values after the transformation.**

the loss functions for one or more experiments along with any choice of hyperparameters. We also set a plausible wide range of values for each model parameter, with some ranges in the log scale, if appropriate. Subsequently, we partition the data of each experiment, allocating 80% for training and the remaining 20% for testing. After completing these steps, we begin the optimization process by using Optuna as the optimization framework, along with the TPE sampling strategy. Within the optimization framework, a set of parameters is sent to a simulator (in the context of this work, that would be Keysight EDA Device Modeling (IC-CAP) [21]). The loss function is evaluated based on the simulator's output for the provided set of model parameters, and an error is returned. Using this information, the optimization framework makes an informed decision to choose the parameters of the subsequent trial. This procedure is repeated for a predetermined number of trials. The model extraction is then judged using the data allocated for testing. If the performance is not satisfactory, we can try a different choice of model, loss function, or hyperparameters to check if any of these settings are more suitable than others. Once we have established the choice of model, loss function, and hyperparameters, we retrain using all available data. This time, however, we further tighten the upper and lower limits on the parameter ranges to be closer to the best set of model parameters identified during the train/test split. We also initiate the optimization process using the best set of parameters we have previously found. Retraining generally requires fewer trials since we already have prior knowledge of where a good set of model parameters can be found. The reasoning behind re-training on the complete dataset is to maximize the model's exposure to all of the available data. This allows the model to learn from the full spectrum of device behaviors and characteristics, which were withheld during the testing phase. It is especially beneficial in semiconductor modeling, where data may be limited, as it can significantly enhance the model's accuracy. This step is also particularly useful when data from a new experiment becomes available from another device fabrication run (which can alter the device characteristics due to process variations) or when the foundry improves the device performance. The steps of our extraction process are summarized in Fig. 1(a). Furthermore, our proposed approach is compared against the conventional extraction flow of the ASM-HEMT model as shown in Fig 1(b), which consists of several manual iterative steps, as outlined in [20].

A. SELECTION OF LOSS FUNCTION

In the context of device modeling, selecting the appropriate loss function is critical to obtaining a reasonable percent error over a range of interest, which ensures a good model fit. Our proposed loss function is chosen carefully to address three significant issues in device modeling. We start with the absolute error (or L1 loss) function as a foundation and then incrementally enhance it to tackle these issues comprehensively, as given by:

$$\mathcal{L}_1(\hat{y}_i, y_i) = |\hat{y}_i - y_i|. \tag{2}$$

First, we want to ensure consistent model performance across different orders of magnitude. For example, we aim to fit the model across a wide range of current values, ranging from small to large (e.g., 10 μA to 100 mA). To achieve this, we perform a log transformation, shifting the focus of the loss function from absolute errors to relative errors. This provides a uniform assessment across different scales of data, resulting in the following equation:

$$\mathcal{L}_{\log}(\hat{y}_i, y_i) = |\log(\hat{y}_i) - \log(y_i)|. \tag{3}$$

Second, we are interested in fitting our model in key operational regions of the device. This implies that we are not interested in fitting the model in less critical regions below some threshold, as spending effort to fit a model to irrelevant values or below the experimental or simulation noise floor will lead to a much worse model. To effectively address this, we transform the loss function once more as follows:

$$u(\hat{y}_i, y_i) = \left| \log\left(1 + \frac{\hat{y}_i}{\epsilon_i}\right) - \log\left(1 + \frac{y_i}{\epsilon_i}\right) \right|. \quad (4)$$

In this second transformation, the hyperparameter ϵ_i denotes a threshold below which the values of $\hat{y}$ and y are considered negligible. This can be exemplified in Fig. 2 by considering the $I - V$ characteristics of a diamond Schottky diode. In Fig. 2, the current values on the left y-axis are represented linearly without the transformation, whereas the right y-axis represents the current values after the transformation. We observe that below the threshold ϵ_i, any current values below 10^{-10} A are effectively excluded from the fitting process (i.e., their contribution is close to zero).

Lastly, we would like a fitting method that is robust to outliers and measurement errors. This is because, in both measurements or simulations, there are some experiments that are simply corrupted, which can occur from hitting instrument compliance at random or when poor accuracy is obtained at higher frequencies, such as in S-parameter measurements. If proper care is not taken, outliers and measurement errors can significantly destroy the quality of the fit. Preferably, it is desired to identify potential outliers and reduce their impact on the fitting process rather than completely excluding them. This can be addressed by incorporating a clipped (non-convex) penalty function on top of equation (4), which is represented mathematically as

$$\mathcal{L}_{\text{clip}}(\hat{y}_i, y_i) = \begin{cases} u^2 & \text{if } |u| \leq \delta_i, \\ \delta_i^2 & \text{if } |u| > \delta_i. \end{cases} \quad (5)$$

Here, u is the transformed error term from (4), which addresses the two previously discussed issues. The hyperparameter δ_i is a predefined threshold that sets the maximum allowable error. This implies that the penalty function puts a fixed cap on $|u|$ larger than δ_i, regardless of size, i.e., we disregard any $|u|$ exceeding δ_i, treating them as outliers or flawed data [22]. Furthermore, to maintain consistency with this methodology, we typically select constant values for the hyperparameters ϵ_i and δ_i for each experiment. For the remainder of this work, unless stated otherwise, the loss function defined in equation (5) will be the primary loss function used for fitting. It incorporates the enhancements and transformations previously detailed to ensure robust and accurate model fitting across various regions of operation.

B. IMPORTANCE OF MODEL ASSESSMENT

When performing model parameter extraction with tens of model parameters, the concept of model assessment becomes essential to ensure the efficacy of our extracted model. This principle is straightforward: the compact model should perform reliably within SPICE simulations under a variety of voltages, currents, and conditions that may differ from our original measurements used to extract the model. We aim for the model to excel in scenarios it has not previously encountered. This is a common problem in Statistics and ML, where the objective is not to match the data used to train these models but to accurately predict new, unseen data [23].

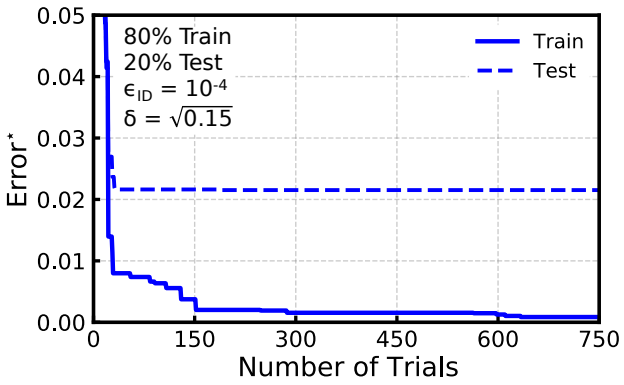


FIGURE 3. Error* versus the number of trials, depicting the minimum error achieved across 750 trials for both train and test curves. In this case, 35 model parameters were optimized to fit 35 measurements, with 80% of the data used for training and the rest for testing.

A simple example illustrating the importance of model assessment can be seen in scenarios where we have 35 model parameters that need to be adjusted based only on 35 measurements. Here, we will use the ASM-HEMT model and 35 $I_D - V_D$ measurements of a 150 nm gate length GaN-on-SiC HEMT, selected randomly. We then split the data into 80% training and the remaining 20% for testing purposes. The results show that the model does an excellent job fitting the training data as the minimum error found to date (Error*) keeps decreasing as the number of trials increases. However, when evaluated against the testing data, the Error* reaches a plateau relatively early. This indicates that the model while improving on the training dataset with more trials, shows little to no improvement on the testing dataset after 30 trials. These results are summarized in Fig. 3.

For these reasons, we adopted the concept of model assessment to judge our model's fit. In all of the fitting examples we will show in the remaining text, we have performed this approach by splitting the data from each experiment into 80% for training and the remaining 20% for testing. The results also show that in most cases, the testing curve for the minimum error found to date is lower than the training curve. As such, this implies that our model generalizes well even for data that it has not seen.

C. DIAMOND SCHOTTKY DIODE FITTING EXAMPLE

Having established the parameter extraction approach and the proposed loss function, we initially focus on applying these methods to a simple two-terminal device, specifically

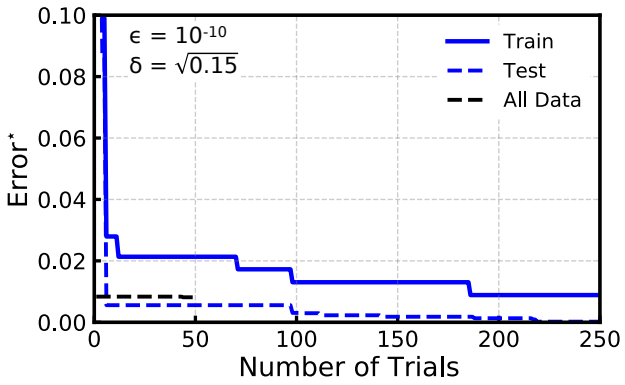


FIGURE 4. **The progression of the minimum Error* observed to date across 250 trials. It displays the train and test curves, as well as the error curve for the entire dataset in the modeling of the diamond Schottky diode over an increasing number of trials. The loss function employed hyperparameter values of $\epsilon = 10^{-10}$ and $\delta = \sqrt{0.15}$.**

a diamond Schottky diode, using the SPICE Diode Model [24]. This demonstration aims to illustrate the effectiveness of our approach in a straightforward setting before tackling the complexities of a more sophisticated compact model. The two-terminal device we are considering is a diamond pseudo-vertical Schottky barrier diode with a diameter of 100 μm. Interested readers may refer to [25] for a more in-depth discussion on the characteristics and performance of these devices, where additional insights into diamond Schottky diodes are presented. The diamond Schottky diode was fabricated on a single crystalline diamond structure consisting of MPCVD-grown p-type epilayers on a type Ib (100) diamond substrate. Additional details of the fabrication and structure of the Schottky diode can be found in [26], where it is referred to as Sample A. Additionally, Mo was used as the Schottky contact metal for the diode measured in this study.

The SPICE diode model is fitted by adjusting three parameters: n (ideality factor), IS (saturation current), and RS (series resistance). The parameter ranges for the SPICE diode model used during the train/test split are provided in Table 1. The model was trained with hyperparameters set to an ϵ of 10^{-10} and a δ of $\sqrt{0.15}$. Having established the loss function and the hyperparameter values, the model was then re-trained for 50 trials to further adjust the model parameters on the entire dataset, starting from the best parameters found during the train/test split. The minimum error found to date is plotted in Fig. 4 as a function of the number of trials for both train and test curves, as well as the error curve for the entire dataset. The $I - V$ characteristics of the simulated data against measurements are displayed in Fig. 5. The fit was then evaluated by considering all current values $> 10^{-10}$ A using the loss function defined in (5), yielding an error of 0.010.

To showcase the efficacy of our proposed loss function against measurement outliers, we intentionally corrupt 50% of the diode's measurements (in practice, this would not occur, but our goal with this example is to show that our fitting is resilient to measurement anomalies). We consider the loss

TABLE 1. **Parameter Ranges in the SPICE Diode Model**

Parameter	Lower Limit	Upper Limit
IS^*	1×10^{-25}	1×10^{-22}
n	0.5	1.5
RS	100	150

*Log-spacing

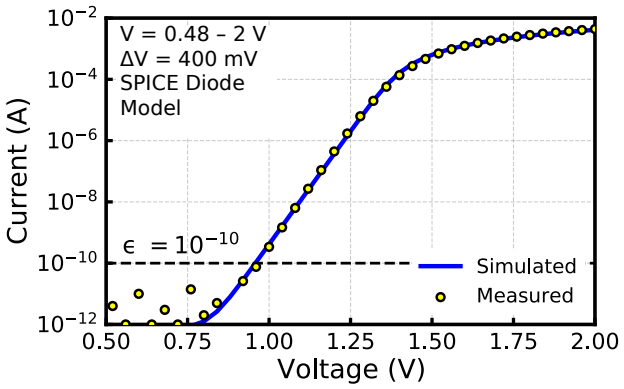


FIGURE 5. **Measured and simulated $I - V$ characteristics of a diamond Schottky diode. The voltage sweep ranged from 0.48 to 2 V at increments of 40 mV, resulting in 39 measurements.**

function defined in (5), which is robust to outliers, as well as a separate non-robust loss function given by

$$\mathcal{L}_{2,u}(\hat{y}_i, y_i) = \left| \log\left(1 + \frac{\hat{y}_i}{\epsilon_i}\right) - \log\left(1 + \frac{y_i}{\epsilon_i}\right) \right|^2. \quad (6)$$

This loss function is similar to (5), but without the addition of the penalty function. Both loss functions are then used to repeat the model extraction procedure with the same hyperparameter values and settings as the previous diode example without data corruption. The minimum error found to date is plotted in Fig. 6(a) as a function of the number of trials for both train and test curves, as well as the error curve when re-training on the entire dataset, for the robust loss function given by (5). Similarly, Fig. 6(b) showcases the train and test curves, along with the error curve when re-training on the entire dataset, for the non-robust function without the penalty function as defined in (6). The $I - V$ characteristics of the simulated data (with corruption) compared to the non-corrupted measurements are displayed in Fig. 7. Despite the data corruption, the robust loss function accurately fitted the model to the measurements, showing an error of 0.011 for currents above 10^{-10} A when compared against the non-corrupted data. Conversely, using the non-robust loss function resulted in a poor fit since it was heavily affected by outliers, yielding an error of 0.756 for currents $> 10^{-10}$ A when compared against the non-corrupted data.

IV. 150-NM GAN-ON-SIC HEMT FITTING EXAMPLE

Having successfully modeled a diode with our proposed approach for parameter extraction, we then shifted our focus to fitting measured DC data of a GaN-on-SiC HEMT to the

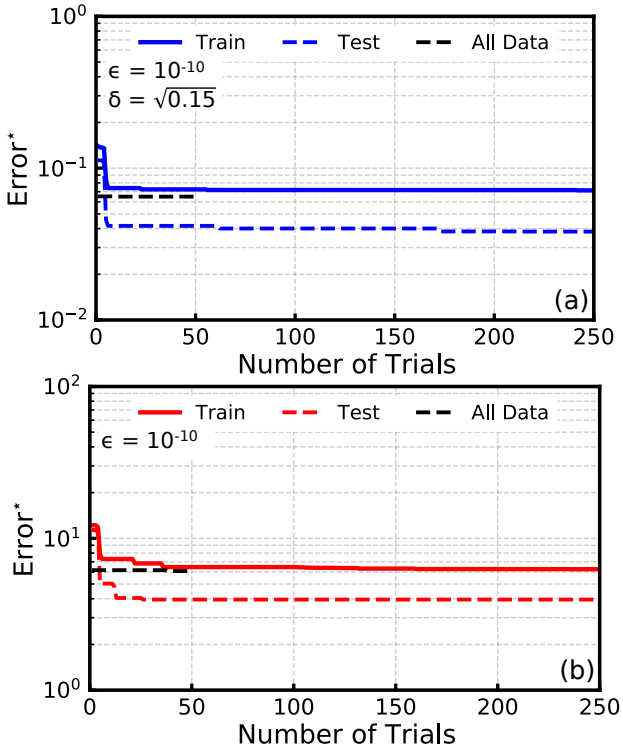


FIGURE 6. **The progression of the minimum Error* observed to date over 250 trials. It showcases the train and test curves, as well as the error curve for the entire dataset in the modeling of the diamond Schottky diode for the (a) robust and (b) non-robust loss functions, with 50% of the data being corrupted. An $\epsilon = 10^{-10}$ was utilized for both loss functions, while a $\delta = \sqrt{0.15}$ was specifically employed for the robust loss function.**

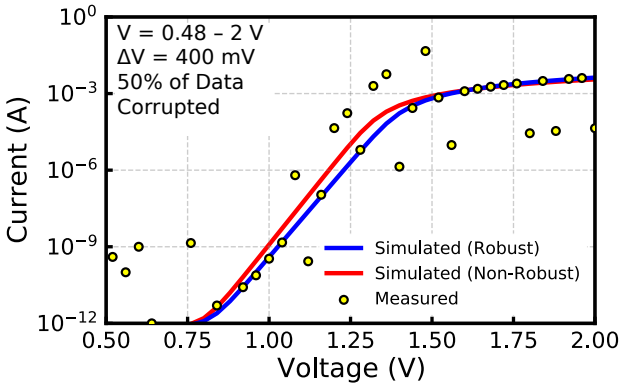


FIGURE 7. **Measured and simulated $I - V$ characteristics of a diamond Schottky diode using robust and non-robust loss functions. Voltage sweep ranged from 0.48 to 2 V in 40 mV steps, totaling 39 measurements. 50% of the measurements were randomly corrupted to evaluate the performance of the loss functions.**

ASM-HEMT DC model [2]. For an in-depth description of the broader implications and technical details of RF GaN HEMT devices and their significance in 5G and beyond-5G wireless communications, interested readers may refer to [27]. The GaN-on-SiC HEMT device we are modeling features a 150-nm gate length and a gate width (W_G) of 4×50 μm. We have previously reported other modeling approaches utilizing the same GaN process in [28], [29]. In this example, we focus on the ASM-HEMT model, an industry-standard compact model for GaN HEMTs. The ASM-HEMT model effectively captures a range of device non-idealities, including self-heating and temperature dependence, mobility degradation, drain-induced-barrier-lowering (DIBL), velocity saturation, and trapping effects [2]. The parameter extraction process for this model begins with extracting DC model parameters, which are fundamental for determining the device's operation as a whole. Considering its importance, it is crucial to pay close attention during this initial stage to obtain an overall good model fitting. The conventional approach to extracting the ASM-HEMT DC model (as described in [20] and [30]) considers splitting the set of DC model parameters into smaller subsets to fit certain regions of operation in the $I - V$ plane through a series of iterative steps. On the other hand, our proposed DFO approach allows us to simultaneously consider 35 relevant parameters in the DC model across a relatively wide range of plausible parameter values and fit them to measured $I - V$ characteristics in a straightforward manner.

In the present work, we used version 101.4.0 of the ASM-HEMT model [20]. To accommodate reverse gate leakage current, we set the *GATEMOD* parameter to 2. The self-heating model was incorporated, whereas the trapping and field-plate models were not. For the device's resistance characteristics, we modeled the bias-dependent access region resistance and source contact resistance by setting *RDSMOD* to 1. Additionally, we enabled a sophisticated gate resistance model by setting *RGATEMOD* to 2 [20]. For our optimization process, we selected 35 model parameters to model the DC characteristics of a GaN-on-SiC HEMT. The particular parameters of the ASM-HEMT model that were modified include *CDSCD, DELTA, ETA0, GDSMIN, IGDDIO, IGSDIO, IMIN, LAMBDA, MEXPACCD, MEXPACCS, NFACTOR, NJGD, NJGS, NS0ACCD, NS0ACCS, RDC, RIGDDIO, RIGSDIO, RNJGD, RNJGS, RSC, RTH0, THESAT, U0, U0ACCD, U0ACCS, UA, UB, UTE, UTED, UTES, VDSCALE, VOFF, VSAT*, and *VSATACCS*.

We followed the steps in our proposed approach, as highlighted in Fig. 1(a). We then fit the ASM-HEMT DC model to the device's drain current (I_D) and transconductance (g_m) characteristics as a function of the gate voltage (V_G) and drain voltage (V_D), corresponding to two different experiments. The voltages of the two experiments were swept from 0 to 20 V (step size of 100 mV) and -3 to -0.1 V (step size of 100 mV) for V_D and V_G, respectively. This results in a total of 6,030 measurements for I_D and 6,030 measurements for g_m. The range of the model parameters is chosen to cover a wide range of plausible values, with the exception of *RTH0*, which was extracted utilizing the technique outlined in [31], and *VOFF*, which can be extracted with relative ease. The parameter ranges for the ASM-HEMT model used during the train/test split are provided in Table 2.

Given that we are dealing with a multi-objective optimiza-

TABLE 2. Parameter Ranges in the ASM-HEMT Model

Parameter	Lower Limit	Upper Limit
CDSCD	1×10^{-3}	150×10^{-3}
ETA0	10×10^{-3}	100×10^{-3}
DELTA	2	5
*GDSMIN**	1×10^{-12}	1×10^{-6}
IGDDIO	7.5	15
IGSDIO	2.5	10
*IMIN**	1×10^{-15}	1×10^{-12}
LAMBDA	100×10^{-6}	1500×10^{-6}
MEXPACCS	1	5
MEXPACCD	1	5
NFACTOR	0.2	0.5
NJGD	2.5	20
NJGS	2.5	20
*NS0ACCD**	5×10^{15}	5×10^{20}
*NS0ACCS**	5×10^{15}	5×10^{20}
RDC	100×10^{-6}	1500×10^{-6}
RIGDDIO	10×10^{-9}	100×10^{-9}
RIGSDIO	10×10^{-9}	100×10^{-9}
RNJGD	15	30
RNJGS	5	15
RSC	100×10^{-6}	1500×10^{-6}
RTH0	31.5	32.5
THESAT	1	4
U0	150×10^{-3}	300×10^{-3}
U0ACCD	50×10^{-3}	250×10^{-3}
U0ACCS	50×10^{-3}	250×10^{-3}
UA	1×10^{-8}	50×10^{-8}
*UB**	1×10^{-21}	1×10^{-18}
UTE	-1	-0.1
UTED	-17.5	-5
UTES	-17.5	-5
VDSCALE	2	6
VOFF	-2.1	-1.9
VSAT	150×10^3	250×10^3
VSATACCS	10×10^3	150×10^3

*Log-spacing

tion problem, as it involves two experiments (i.e., I_D and g_m), we apply a simple scalarizer to transform the multi-objective problem into a single-objective problem. This weighted sum scalarizer is given by

$$\phi(u) = w^T v, \tag{7}$$

where $w \in \mathbb{R}^n$ is a set of weights, with $w_i \geq 0$, and v_i are the various loss functions or objectives. The model was initially trained for 10,000 trials after conducting a test/train split, with hyperparameters set to an ϵ of 10^{-10}, a δ of $\sqrt{0.15}$, and weights $w_1 = 0.50$ and $w_2 = 0.50$, corresponding to the weighted sums of the errors for I_D and g_m, respectively. After establishing the loss function and setting the hyperparameters, the model was re-trained for an additional 2,500 trials on the entire dataset, beginning with the best parameters identified during the train/test split step. The minimum error found to date is plotted in Fig. 8 as a function of the number of trials, showing both the train and test curves as well as the error curve for the entire dataset. We note that the best set of model parameters was found in less than 6,000 trials, resulting in an excellent fit with an error of 1.25e-3 for current values $> 10^{-4}$ A and an error of 2.17e-3 for transconductance

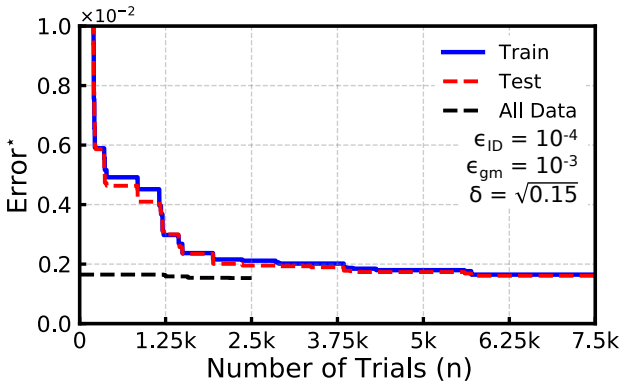


FIGURE 8. The progression of the minimum Error* observed to date for an increasing number of trials. It showcases the train and test curves and the error curve for the entire dataset in the modeling of the GaN-on-SiC HEMT. The loss functions employed hyperparameter values of $\epsilon = 10^{-4}$ for I_D and $\epsilon = 10^{-3}$ for g_m. A $\delta = \sqrt{0.15}$ was used for both experiments.

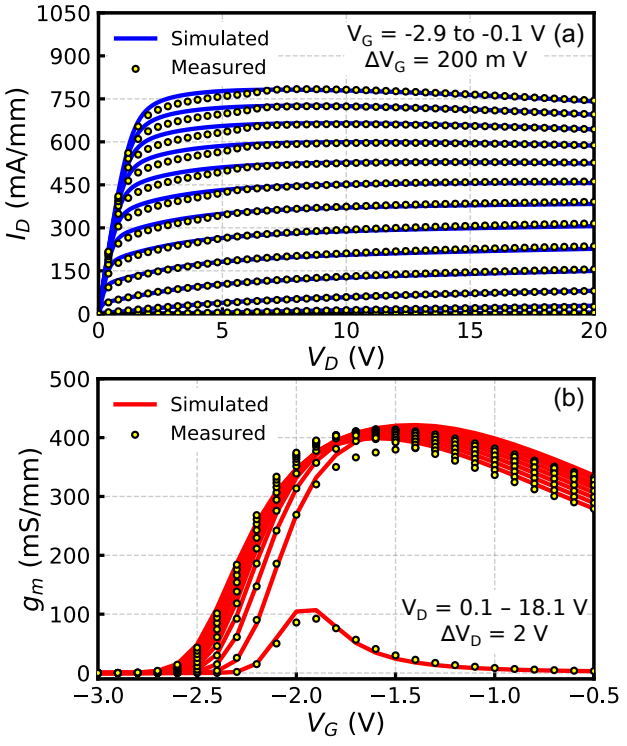


FIGURE 9. Measured and simulated (a) I_D and (b) g_m characteristics of the GaN HEMT. For (a), the drain voltage V_D was swept from 0 to 20 V (step size of 100 mV), and V_G was swept from -2.9 to -0.1 V (step size of 200 mV). For (b), the gate voltage V_G was swept from -3 to -0.5 V (step size of 100 mV), and V_D was swept from 0.1 to 18.1 V (step size of 2 V).

values $> 10^{-3}$ S. The measured and simulated I_D and g_m characteristics are shown in Fig. 9.

Just as in the diode example, we assess robustness by intentionally corrupting 25% of the measurements randomly. In this instance, we introduce outliers by transforming the DC characteristics from the linear domain to the logarithmic domain, corrupting the data, and then reverting back to the

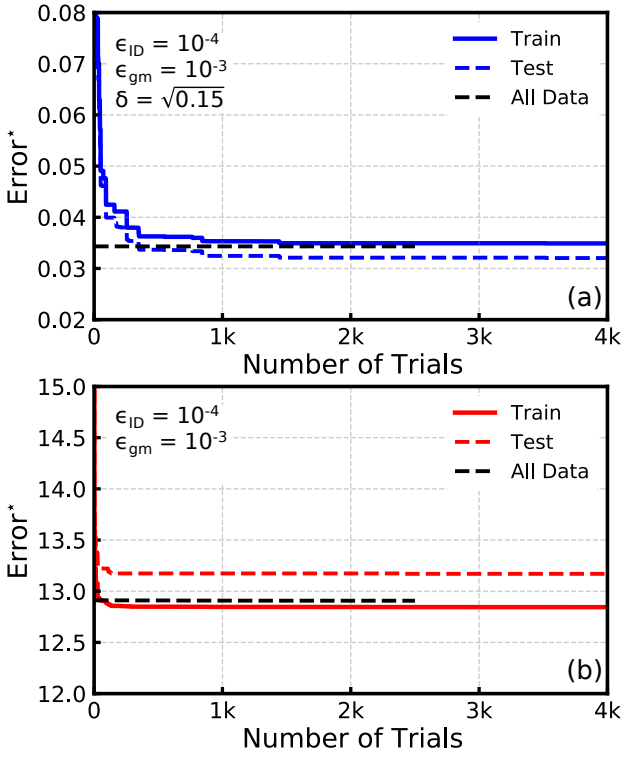


FIGURE 10. The progression of the minimum Error* observed to date over an increasing number of trials. It showcases the train and test curves, as well as the error curve for the entire dataset in the modeling of the GaN-on-SiC HEMT for the (a) robust and (b) non-robust loss functions, with 25% of the data being corrupted. The loss functions employed hyperparameter values of $\epsilon = 10^{-4}$ for I_D and $\epsilon = 10^{-3}$ for g_m. For the robust loss function, a $\delta = \sqrt{0.15}$ was used for both experiments.

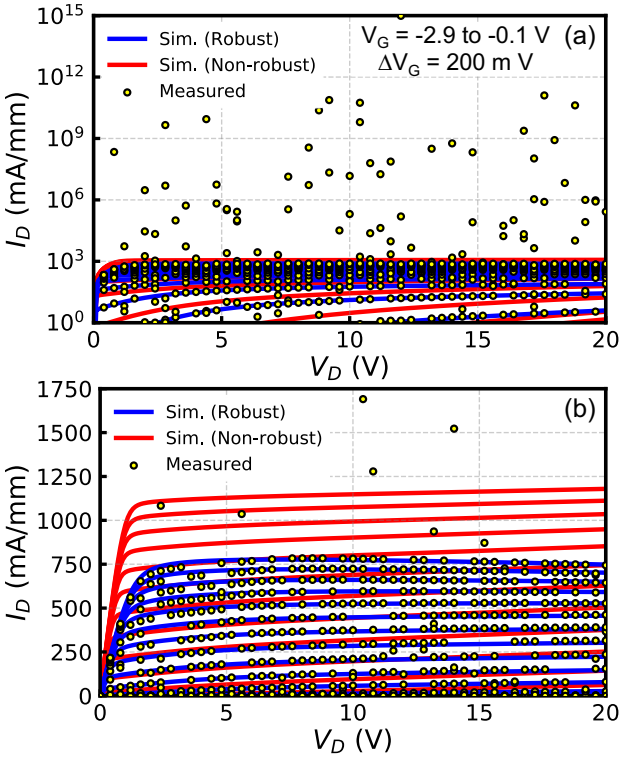


FIGURE 11. Measured and simulated output (I_D) characteristics of the GaN-on-SiC HEMT in the (a) log and (b) linear regime. The V_D was swept from 0 to 20 V (step size of mV), and V_G was swept from -2.9 to -0.1 V (step size of 200 mV). 25% of the measurements were corrupted at random based on a normal (Gaussian) distribution with $\mu = 0$ and $\sigma = 10$.

linear domain. These outliers follow a normal (Gaussian) distribution with a mean (μ) of 0 and a standard deviation (σ) of 10. The logarithmic and linear representations of the measured and simulated I_D characteristics are displayed in Fig. 11. We follow the same procedure as in the diode example, where we use the loss functions from (5) and (6) to test their robustness against measurement outliers. The minimum error found to date for both loss functions is plotted in Figs. 10(a) and (b) as a function of the number of trials. This includes both the train and test curves, as well as the error curve when re-training (for 2,500 trials) on the entire dataset. The measured and simulated corrupted I_D and g_m characteristics are also displayed in Fig. 11. Once again, this shows that our method is robust against outliers, as our extracted model using the robust loss function as defined in (5) achieved an error of 1.27e-3 for currents $> 10^{-4}$ A and an error of 2.53e-3 for transconductance values $> 10^{-3}$ S when compared against the non-corrupted data. On the other hand, the non-robust loss function as defined in (6) resulted in a poor fit, with an error of 0.325 for currents $> 10^{-4}$ A and an error of 0.238 for transconductance values $> 10^{-3}$ S when compared against the non-corrupted data.

V. CONCLUSION

In this paper, we have addressed the problem of model parameter extraction via derivative-free optimization. We propose using a loss function that focuses on relative errors to ensure consistent model performance across varying magnitudes, prioritizes the accuracy above a certain measurement or simulation threshold, and is robust against outliers and measurement. We then considered two examples, starting with the modeling of a diamond Schottky diode, to showcase how we address these three critical issues. Our focus is then shifted to extracting the ASM-HEMT DC model in a 150-nm GaN-on-SiC HEMT by simultaneously extracting 35 model parameters. Our proposed approach yields good results in a fraction of the simulations that are required by other approaches in the literature. For instance, the work in [8] demands 120,000 simulation runs for training and testing while considering only 10 ASM-HEMT model parameters. This deep learning approach takes a significant amount of time to gather the required training and testing data and could require an even more substantial number of simulations if additional model parameters are considered. On the other hand, our approach achieved a good fit in a fraction of that amount (i.e., <5%), demonstrating the usefulness of DFO in device modeling. Lastly, to foster further exploration of our approach within the device modeling community, we will make our code open-

source, which can be found in our GitHub repository.[1]

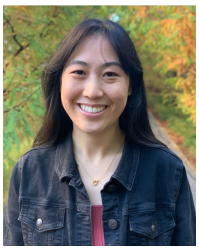
REFERENCES

[1] "BSIM-CMG 111.2.1," https://bsim.berkeley.edu/models/bsimcmg/, accessed: 2024-02-24.
[2] S. Khandelwal, *Advanced SPICE Model for GaN HEMTs (ASM-HEMT): A New Industry-Standard Compact Model for GaN-based Power and RF Circuit Design*. Cham, Switzerland: Springer, 2022.
[3] P. Kuthe, M. Müller, and M. Schröter, "VerilogAE: An Open Source Verilog-A Compiler for Compact Model Parameter Extraction," *IEEE J. Electron Devices Soc.*, vol. 8, pp. 1416–1423, 2020.
[4] K. Levenberg, "A method for the solution of certain problems in least squares," *Quart. Appl. Math.*, vol. 5, p. 164–168, 1944.
[5] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *J. Soc. Ind. Appl. Math.*, vol. 11, no. 2, pp. 431–441, Jun 1963.
[6] L. Duc-Hung, P. Cong-Kha, N. Thi Thien Trang, and B. Trong Tu, "Parameter extraction and optimization using Levenberg–Marquardt algorithm," in *Proc. 4th Int. Conf. Commun. Electron. (ICCE)*, Aug 2012, pp. 434–437.
[7] M. Y. Kao, F. Chavez, S. Khandelwal, and C. Hu, "Deep Learning-Based BSIM-CMG Parameter Extraction for 10-nm FinFET," *IEEE Trans. Electron Devices*, vol. 69, no. 8, pp. 4765–4768, Aug. 2022.
[8] F. Chavez, D. T. Davis, N. C. Miller, and S. Khandelwal, "Deep Learning-Based ASM-HEMT I-V Parameter Extraction," *IEEE Electron Device Lett.*, vol. 43, no. 10, pp. 1633–1636, Oct. 2022.
[9] F. Chavez and S. Khandelwal, "Deep Learning-Based ASM-HEMT High Frequency Parameter Extraction," in *IEEE Wirel. Micro. Technol. Conf. (WAMICON)*, 2023, pp. 41–44.
[10] A. Arias-Purdue, E. Lam, J. Tao, E. O'Malley, and J. F. Buckwalter, "Artificial Neural Networks for GaN HEMT Model Extraction in D-band Using Sparse Data," in *IEEE Radio Freq. Integr. Circuits Symp. (RFIC)*, San Diego, CA, USA, 2023, pp. 65–68.
[11] F. Chavez and S. Khandelwal, "Machine Learning-Based Large-Signal Parameter Extraction for ASM-HEMT," *IEEE Microwave Wirel. Technol. Lett.*, vol. 34, no. 2, pp. 147–150, Feb. 2024.
[12] J. Chen, F. Chavez, C.-T. Tung, S. Khandelwal, and C. Hu, "A single neural network global I-V and C-V parameter extractor for BSIM-CMG compact model," *Solid-State Electronics*, vol. 216, p. 108898, 2024.
[13] R. P. Martinez, S. Boyd, and S. Chowdhury, "Robust Pareto Design of GaN HEMTs for Millimeter-Wave Applications," 2024.
[14] P. Koch, O. Golovidov, S. Gardner, B. Wujek, J. Griffin, and Y. Xu, "Autotune: A Derivative-Free Optimization Framework for Hyperparameter Tuning," in *Proc. ACM SIGKDD Conf. Knowl. Discov. Data Mining*, 2018, p. 443–452.
[15] G. Maher, S. Boyd, M. Kochenderfer, C. Matache, A. Ulitsky, S. Yukhymuk, and L. Kopman, "A Light-Weight Multi-Objective Asynchronous Hyper-Parameter Optimizer," *arXiv:2202.07735*, 2022.
[16] J. Bergstra, B. Komer, C. Eliasmith, D. Yamins, and D. D. Cox, "Hyperopt: A python library for model selection and hyperparameter optimization," *Comput. Sci. Discovery*, vol. 8, no. 1, p. 014008, 2015.
[17] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*. ACM, 2019, pp. 2623–2631.
[18] O. Kramer, D. E. Ciaurri, and S. Koziel, "Derivative-free optimization," in *Computational Optimization, Methods and Algorithms*. Berlin, Germany: Springer, 2011, pp. 61–83.
[19] S. Watanabe, "Tree-Structured Parzen Estimator: Understanding Its Algorithm Components and Their Roles for Better Empirical Performance," 2023.
[20] S. Khandelwal *et al.*, *ASM-HEMT 101.4.0: Advanced SPICE Model for HEMTs*, Technical Manual, 2023.
[21] Keysight Technologies, "Keysight EDA Device Modeling (IC-CAP)," http://www.keysight.com/find/device-modeling, 2024, accessed: May 10, 2024.
[22] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
[23] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY, USA: Springer, 2009.
[24] P. Antognetti and G. Massobrio, *Semiconductor Device Modeling with SPICE*. New York: McGraw-Hill, 1988.
[25] K. Woo, Z. Bian, M. Noshin, R. P. Martinez, M. Malakoutian, B. Shankar, and S. Chowdhury, "From wide to ultrawide-bandgap semiconductors for high power and high frequency electronic devices," *J. Phys.: Mater*, vol. 7, no. 2, 2024.
[26] K. Woo, M. Malakoutian, D. Saraswat, Z. Bian, A. Hardy, M. Muehle, T. A. Grotjohn, and S. Chowdhury, "Control of Schottky barrier height in diamond using UV-generated ozone and its effect on barrier inhomogeneity and temperature dependent properties," *Diam. Relat. Mater.*, p. 111059, 2024.
[27] H. Lu *et al.*, "A Review of GaN RF Devices and Power Amplifiers for 5G Communication Applications," *Fundam. Res.*, 2023.
[28] R. P. Martinez, M. Iwamoto, J. Xu, P. Pahl, and S. Chowdhury, "Benchmarking Measurement-Based Large-Signal FET Models for GaN HEMT Devices," in *Proc. IEEE Radio Freq. Integr. Circuits Symp.*, Jun. 2023, pp. 69–72.
[29] R. P. Martinez *et al.*, "Assessment and Comparison of Measurement-Based Large-Signal FET Models for GaN HEMTs," *IEEE Trans. Microw. Theory Techn.*, vol. 72, no. 5, pp. 2692–2703, May 2024.
[30] S. A. Ahsan, S. Ghosh, S. Khandelwal, and Y. S. Chauhan, "Physics-Based Multi-Bias RF Large-Signal GaN HEMT Modeling and Parameter Extraction Flow," *IEEE J. Electron Devices Soc.*, vol. 5, no. 5, pp. 310–319, Sep. 2017.
[31] D. E. Root, J. Xu, and M. Iwamoto, "Thermal Resistance Formulation and Analysis of III-V FETs Based on DC Electrical Data," in *Proc. IEEE BiCMOS Compound Semicond. Integr. Circuits Technol. Symp. (BCICTS)*, Dec. 2021, pp. 1–4.

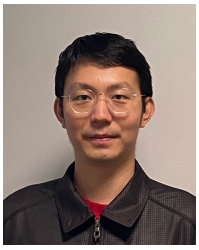
[1] https://github.com/rafapm/dfo_parameter_extraction

RAFAEL PEREZ MARTINEZ (Graduate Student Member, IEEE) received the B.S. degree (Hons.) in electrical engineering and the M.S. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2016 and 2019, respectively, and the M.S. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 2023, where he is currently pursuing the Ph.D. degree in electrical engineering. From summer 2022 to 2023, he was an Intern at Keysight Technologies Inc., Santa Rosa, CA, USA, where he was responsible for GaN device modeling, characterization, and circuit design. His current research interests include III–V device modeling, optimization, and characterization for RF/mm-wave applications.

MASAYA IWAMOTO (Member, IEEE) received the B.S. degree in electrical engineering from Cornell University, Ithaca, NY, in 1997, and the M.S. and Ph.D. degrees in electrical engineering from the University of California at San Diego, La Jolla, in 1999 and 2003, respectively. He has been with Keysight Technologies (formerly Agilent Technologies) in Santa Rosa, California, since 2003, where he is involved in the area of GaAs and InP-based IC technology development. He is the key developer of the ADSHBT Model (formerly AgilentHBT Model), which is available in several of Keysight's circuit design software products and has been used in the design of GaAs and InP HBT ICs for Keysight's microwave/mm-wave instrumentation. Currently, he is working on the reliability of HBTs and HEMTs in various III-V material systems (InP, GaAs, GaN) where his research interest is to apply advanced modeling and characterization techniques in the evaluation of these devices.



KELLY WOO (Student Member, IEEE) received the B.S. degree in electrical engineering from the California Institute of Technology, Pasadena, California, USA, in 2018, and the M.S. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 2020, where she is currently pursuing the Ph.D. degree in electrical engineering. Her current research interests include fabricating various high-power electronic devices based on diamond technology and other wide-bandgap materials.



ZHENGLIANG BIAN (Student Member, IEEE) received the B.S. degree in engineering mechanics from Tsinghua University, Beijing, China, in 2019, and an M.S. degree in electrical engineering from Stanford University, Stanford, California, USA, in 2022, where he is currently pursuing the Ph.D. degree in mechanical engineering. His research interests include GaN-based high-power and high-frequency devices and novel thermal management methods for these devices.

ROBERTO TINTI received the Ph.D. degree in electrical engineering from Delft Technical University, The Netherlands, in 1999. Since joining Keysight Technologies (formerly HP) in 1999, he has been involved in the area of advanced device characterization, including the development of early pulsed and 1/f Noise measurement systems. He contributed significantly to several areas of the IC-CAP device modeling platform, such as optimization and GaN modeling extraction flows. Recently, he became the Product Owner of Keysight's device modeling software products. Currently, he is focusing on advancing Keysight EDA design solutions to address current and future challenges in device modeling.

STEPHEN BOYD (Life Fellow, IEEE) received the A.B. degree in mathematics from Harvard University, Cambridge, MA, USA, in 1980, and the Ph.D. degree in electrical engineering and computer science from the University of California, Berkeley, CA, USA, in 1985. He is currently the Samsung Professor in engineering, and a Professor in electrical engineering with Stanford University, Stanford, CA, USA. He is a member of U.S. National Academy of Engineering (NAE), a foreign member of the Chinese Academy of Engineering (CAE), and a foreign member of the National Academy of Engineering of Korea (NAEK). His current research focuses on convex optimization applications in control, signal processing, machine learning, and finance.

SRABANTI CHOWDHURY (Fellow, IEEE) is an Associate Professor of Electrical Engineering, Senior fellow of Precourt Institute for Energy and Materials Science and Engineering (by courtesy) at Stanford University. She received the M.S. and Ph.D. degrees in electrical and computer engineering from the University of California, Santa Barbara, Santa Barbara, CA, in 2008 and 2010, respectively, where she demonstrated the first vertical power-switching transistor in GaN, known as CAVET. She specializes in wideband gap and ultra-wide bandgap materials and device engineering, with a focus on energy-efficient system architecture and thermal management. Among her notable recognitions are the 2023 Technical Excellence Award from SRC for her work on diamond integration with GaN and SiC, the 2020 Alfred P. Sloan Fellowship in Physics, and the 2016 Young Scientist Award at the International Symposium of Compound Semiconductors (ISCS) for developing Vertical GaN transistors. She actively serves on various IEEE committees, including EDS, VLSI, and IEDM.