IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 19, 2019, accepted January 23, 2020, date of publication February 3, 2020, date of current version February 10, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2970966

Powernet: SOI Lateral Power Device Breakdown Prediction With Deep Neural Networks

JING CHEN[1,2,3], (Student Member, IEEE), MOHAMED BAKER ALAWIEH[3], (Student Member, IEEE), YIBO LIN[4], (Member, IEEE), MAOLIN ZHANG[1,2], (Student Member, IEEE), JUN ZHANG[1,2], YUFENG GUO[1,2], (Member, IEEE), AND DAVID Z. PAN[3], (Fellow, IEEE)

[1]College of Electronic and Optical Engineering and the College of Microelectronics, Nanjing University of Posts and Telecommunications, Nanjing 210046, China
[2]National and Local Joint Engineering Laboratory for RF Integration and Micro-packaging Technologies, Nanjing University of Posts and Telecommunications, Nanjing 210046, China
[3]Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712, USA
[4]Center for Energy-Efficient Computing and Applications (CECA), School of EECS, Peking University, Beijing 100871, China

Corresponding author: Yufeng Guo (yfguo@njupt.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61574081 and Grant 61874059, in part by the Graduate Students Research Innovation Program of Jiangsu Province under Grant 46106CX18030, and in part by the Chinese Scholarship Council under Grant 201908320339.

ABSTRACT The breakdown performance is a critical metric for power device design. This paper explores the feasibility of efficiently predicting the breakdown performance of silicon on insulator (SOI) lateral power device using multi-layer neural networks as an alternative to expensive technology computer-aided design (TCAD) simulation. In this work, we propose the first breakdown performance prediction framework, *PowerNet*, for SOI lateral power devices, based on deep learning methods. The framework can provide breakdown location prediction and breakdown voltage (BV) prediction by utilizing a two-stage machine learning method. In addition, it demonstrates 97.67% accuracy on breakdown location prediction and less than 4% average error on the BV prediction compared with TCAD simulation. The proposed method can be used to measure changes in performance caused by random variability in structural parameters during manufacturing process, allowing designers to avoid unstable structural parameters and enhance design robustness. More importantly, it can significantly reduce the computational cost when compared with the TCAD simulation. We believe the proposed machine learning technique can significantly speedup the design space exploration for power devices, eventually reducing the overall product-to-market time.

INDEX TERMS Breakdown location, breakdown voltage, deep neural networks (DNN), Gaussian process regression (GPR), SOI lateral power device.

I. INTRODUCTIONS

Silicon on insulator (SOI) lateral power devices have been widely adopted in power integrated circuits due to their high breakdown voltage (BV), excellent isolation, and low power consumption [1]–[3]. Analysis of breakdown performance has always been a hot issue for researchers [4]–[7]. During the manufacturing process, random deviations in the expected structural parameters may result in drastic changes in breakdown performance. Thus, the early prediction of a device breakdown performance stands out as prominent ramification in the device design closure. In literature, many approaches have been proposed to assess the breakdown performance.

The conventional approach relies on device simulation using tools such as Medici [4], Silvaco [5], Sentaurus [6] and so on. Chen *et al.* [4] measured and improved the breakdown performance of a lateral double-diffusion metal-oxide-semiconductor field-effect transistor (MOSFET) with n^+ floating island by using Medici. Zeng *et al.* [5] adopted Silvaco to obtain the breakdown voltage of lateral MOSFET. These simulation tools acquire the breakdown performance by solving the physical equations such as Poisson equation and the current continuity of electrons and holes at the set mesh point. When it comes to high breakdown voltage,

The associate editor coordinating the review of this manuscript and approving it for publication was Liang-Bi Chen.

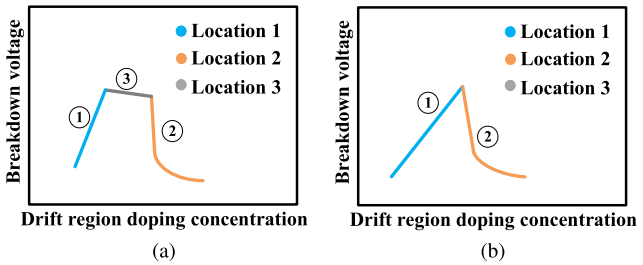


FIGURE 1. The variation of breakdown voltage with drift region doping concentration, (a) when the breakdown location changes from N^+N^- junction to P^+N^- junction with drift region doping concentration, it will also occur in the body of device, i.e. location 3. (b) the breakdown location only occurs at the N^+N^- junction and P^+N^- junction [10], [22], [23].

these methods suffer from exorbitant computational overhead. Moreover, the numerical convergence of these physical functions is sensitive to mesh settings, resulting in long simulation time.

Meanwhile, models based on analytical derivations [8]–[12] are proposed for power devices by solving the Poisson equation with boundary conditions. However, these models are ill-equipped to handle complicated cases with a large number of structural parameters. Jun *et al.* [10] proposed a novel one-dimensional breakdown voltage model of SOI lateral power device. Later, Yao *et al.* [9] presented a 3D analytical model for SOI lateral power device. While these models are computationally efficient, they only consider a small number of the structural parameters in the analysis of breakdown performance.

Recently, machine learning techniques have been widely adopted for different applications in the field of electronic design automation [13]–[18]. It has been widely used for its capability in learning hidden correlations between the electronic parameters and the electric performance with high accuracy and efficiency [19]–[21]. This capability is very promising for early-stage design space exploration to determine a robust device structure in a short design period. In literature, Zeng *et al.* [19] presented an insulated-gate bipolar transistor (IGBT) current measuring method by using artificial neural network and K-means clustering, they achieved 3% average error in the prediction of IGBT current. Carrillo-Nuñez *et al.* [21] proposed the prediction of the effect of statistical variability in Si junctionless nanowire transistors using a multi-layer neural network (NN), which can greatly save computational cost. These results have demonstrated the potential of machine learning approaches in device performance predictions.

Despite the importance of breakdown performance evaluation for power device, the breakdown performance prediction has not yet been tackled using machine learning techniques. In this paper, we propose *PowerNet*, as the first efficient design methodology for power devices with a machine learning based breakdown performance prediction framework. By observing that the breakdown voltage changes rapidly with the structural parameters [10], [22], [23], such as drift region doping concentration N_d as shown in Fig. 1, it is hard to capture the trend with a single model like previous works [19], [20]. We also observe in the figure that the rapid changes of the breakdown voltage come together with the switch of breakdown locations, e.g., from location 1 to location 2, or from location 3 to location 2. Therefore, we propose a novel two-stage modeling technique for *PowerNet*.

In the first stage, the breakdown location is predicted by a two hidden layer neural network. Then, the breakdown voltage is predicted using a deeper location-specific predictive neural network. With this approach, the models in the second stage are location-specific, and hence, are more accurate than a single generalized model for all locations.

The main contributions are summarized as follows.

- We propose *PowerNet* as a novel breakdown performance prediction framework, for the first time, using a two-stage modeling technique, which can achieve a fast prediction and overcome the convergence problem compared with tool simulation.
- We propose a first stage classification model to predict the breakdown location followed by a second stage location-specific regression to predict breakdown voltage for SOI lateral power device.
- The experimental results show that the accuracy of breakdown location prediction is about 97.67% and the average error of breakdown voltage prediction is less than 4% compared with the results from technology computer-aided design (TCAD) simulator [24].

The rest of the paper is organized as follows. Section II reviews the basic background for breakdown performance of SOI lateral power device and introduces the problem formulation. Section III describes detailed algorithms of the proposed method, including first stage deep neural networks (DNN) for the breakdown location prediction and second stage DNN or Gaussian process regression (GPR) for the BV prediction. Then, Section IV validates our proposed methods with experimental results. Finally, Section V concludes the whole study.

II. PRELIMINARIES

In this section, we first introduce the SOI lateral power device structure we adopt and the factors which have an influence on the breakdown performance. Then, the background about the breakdown performance is analyzed. Next, we give the problem formulation for the breakdown performance measuring.

A. LATERAL POWER DEVICE STRUCTURE

There are many types of SOI lateral power devices such as lateral power diodes, laterally diffused MOSFET (LDMOS), IGBT, etc. However, the basic structure of the breakdown of all lateral power devices is a diode. Therefore, we use a lateral power diode structure as an example. Fig. 2a shows the schematic diagram of the SOI lateral power diode structure. The BV can be obtained by applying a voltage to the cathode (N^+) until the device breaks down while the anode (P^+) is grounded. In the breakdown process, a set of structural parameters of the drift region has a significant influence on the breakdown performance. These parameters, summarized

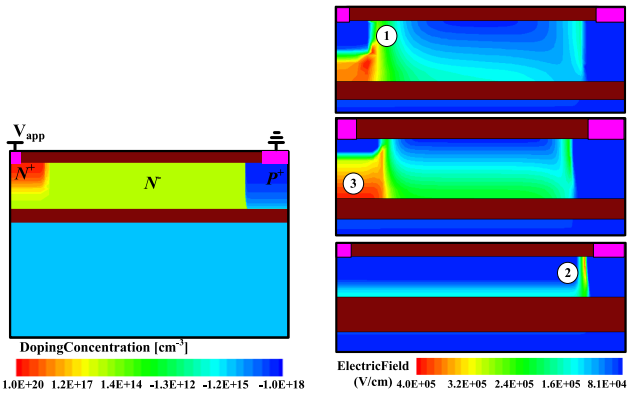


FIGURE 2. Breakdown review of an SOI lateral power device, (a) device structure; (b) electric field distribution with the breakdown location.

TABLE 1. Structural parameters in the drift region.

Structural Parameters	Name	Range
t_{ox}	Buried Oxide Layer Thickness	[0.4,6.5]
N_d	Drift Region Doping Concentration	[1e15,2e16]
t_s	Top Layer Silicon Thickness	[0.4,9]
L	Drift Region Channel Length	[10,135]
l_n	Cathode N^+ Length	[2,14]
N_n	Cathode N^+ Doping Concentration	[1e20,2e21]
t_n	Cathode Junction Depth	[t_s/3,t_s/1.25]
l_p	Anode P^+ Length	[3,15]
P_p	Anode P^+ Doping Concentration	[1e18,2e19]
t_p	Anode Junction Depth	[t_s/1.5,t_s]
r	Radius of Curvature	[0,7]

in Table 1, are used for predicting the BV in our proposed *PowerNet* approach.

B. BREAKDOWN ANALYSIS

With the change in the structural parameters, the breakdown location and BV change correspondingly. Fig. 2b shows the breakdown location corresponding to Fig. 1. With the change of structural parameters, the breakdown voltage will change correspondingly. As the breakdown location moves, the trend of the magnitude of the BV also changes. Primarily, it should be noticed that when the breakdown location is shifted from location 1 or location 3 to location 2, the BV will decrease dramatically. Hence, it is of importance to account for such correlation between the breakdown location and the BV trend when modeling the breakdown behavior. With this in mind, our proposed *PowerNet* comprises two stages where the location of the breakdown is first predicted in stage 1, then a model trained exclusively to estimate BV at the predicted location is used.

C. PROBLEM FORMULATION

The breakdown performance prediction problem can be defined as follows:

Problem 1 (Breakdown Performance Prediction Problem): The goal is to train learning models using training data labeled with specific breakdown information so that the correct breakdown performance can be given using the model when applying another testing data.

Considering the sudden change of breakdown voltage with the change of breakdown position, the learning model is mainly composed of two parts: one is breakdown location prediction, and the other is BV prediction.

Since the breakdown only occurs at the three locations as described in Section II-B, the breakdown location prediction can be analogous to a three-classification problem where each breakdown position corresponds to a category. Therefore, the problem can be defined as follows:

Subproblem 1 (Breakdown Location Prediction Problem): The goal is to train a model using training data labeled with specific locations 1, 2 and 3 so that the correct breakdown locations can be given using the model when applying another testing data.

Besides, since the BV prediction is a regression problem, it can be defined as follows:

Subproblem 2 (Breakdown Voltage Prediction Problem): The goal is to train three models using the three kinds of data labeled with specific breakdown voltage to achieve accurate prediction of the breakdown voltage using the data with known breakdown location.

III. ALGORITHM FOR BREAKDOWN PERFORMANCE MEASURING

In this section, we introduce the detailed components of the proposed *PowerNet* including the DNN model for breakdown location prediction, followed by the BV prediction step where two models with different features, namely DNN and GPR, are presented.

A. OVERALL FLOW FOR POWERNET

In some scenarios, the breakdown voltage changes drastically even with the slightest change in the structural parameters. In other words, the relation between the breakdown voltage and the structural parameters is ''ill-conditioned''. Hence, it is evidently impossible for a single model to globally capture the breakdown behavior; a fact that necessities adopting a local modeling scheme where different BV trends are captured by different models. An important observation to build upon here is that the irregular changes in the BV are usually associated with a switch in the breakdown location. In other words, the BV trend is highly dependent on the breakdown location and can vary significantly when the location changes. Hence, we propose *PowerNet* as a two-stage BV modeling scheme, where the first stage predicts the breakdown location and location-specific models are trained to predict BV at each breakdown location in the second stage.

Fig. 3 shows the overall flow of the two-stage *PowerNet* for breakdown performance evaluation. It comprises two main steps in each of the training and usage phases. At the first step in the training phase, a classifier is trained to predict the breakdown location using the entire training dataset. Next, three regression models are trained to predict the BV at three different locations. For a model at a particular location, only the samples in the training dataset that correspond to that location are used in the training. At the end of the training phase, a two-stage prediction scheme with

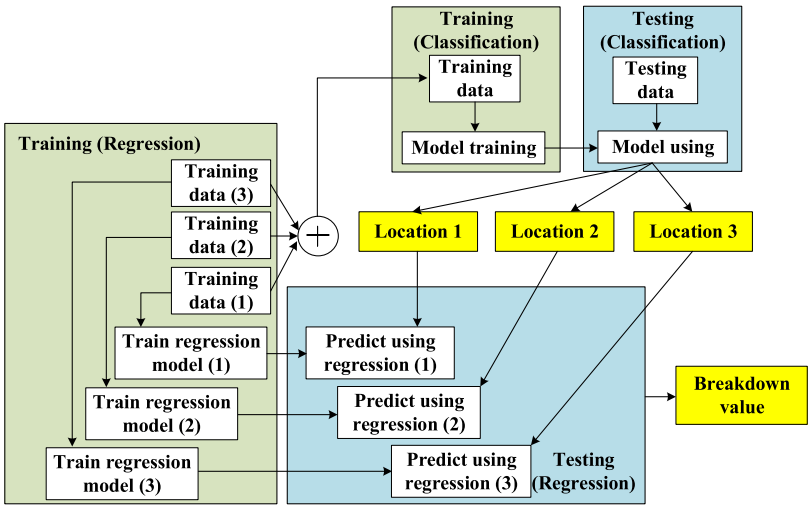


FIGURE 3. The overall flow of the two-stage *PowerNet* for breakdown performance evaluation, in which the DNN is used for classification of breakdown location prediction, and the DNN/GPR is employed for regression of BV prediction.

a global classifier and location-specific regression models are available.

In the usage phase, the process flows in a similar fashion. For each sample in the testing dataset, the classifier is first used to predict the breakdown location. Then, based on the predicted location, the corresponding regression model is used to predict the BV. In *PowerNet*, we propose using DNN as the classification model, whereas for the regression, we present two regression alternatives: (i) DNN and (ii) GPR. Further details about these models will be presented in the following Sections.

B. DEEP NEURAL NETWORK FOR BREAKDOWN LOCATION PREDICTION

A DNN can be utilized to approximate a mathematical model with complex functions by mimicking the biological neuronal system composed of neurons and synapses [25], [26]. Practically, a DNN model is trained using a set of labeled instances to learn a prediction task.

In general, the network architecture is dependent on the number of features and the complexity of the problem at hand. Fig. 4 shows the DNN architecture used for breakdown location prediction of SOI lateral power device (i.e., the classification model). Its structure is based on three connected components: an input layer, a set of two hidden layers and an output layer.

The key idea of the DNN mainly consists of two propagation paths, the information forward and error backward. The forward propagation of information is used to calculate the prediction result by using the weights and bias terms of the inter-layer connections. Then these weights and bias terms are updated based on the error between the prediction result and the actual result in a backward propagation scheme. Backpropagation is the standard algorithm used to train the DNN models [27], by updating the parameters of the network to reduce the loss function. During training, the process of alternating propagation between information forward and error backward is repeated until convergence.

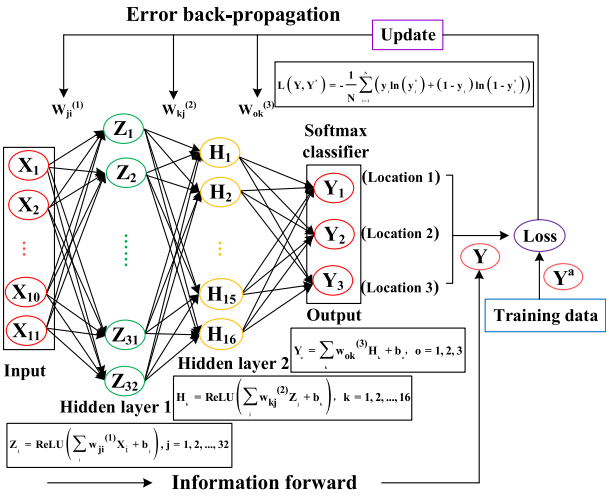


FIGURE 4. Backpropagation DNN architecture for breakdown location prediction. The three components of it are the input layer, two hidden layers, and the output layer for the information forward and error backward during the model training process.

In fact, designing the network architecture is an engineering process that entails multiple trials and adjustments. The major decisions here are concerning the number of hidden layers and that of the neurons in each of the hidden layers. In our study, we build a DNN with two hidden layers to model classification task. Note that the specific configurations of the hidden layers and the corresponding neurons for the breakdown location prediction model are shown in Fig. 4.

According to the DNN structure, the model can be expressed mathematically as follows:

$$\begin{cases} X = (t_{ox}, N_d, t_s, L, l_n, N_n, l_p, P_p, t_n, t_p, r), \\ Z_j = \mathrm{ReLU}(\sum_i w_{ji}^{(1)} X_i + b_j), j = \{1, 2, \ldots, 32\}, \\ H_k = \mathrm{ReLU}(\sum_j w_{kj}^{(2)} Z_j + b_k), k = \{1, 2, \ldots, 16\}, \\ Y_o = \sum_k w_{ok}^{(3)} H_k + b_o, o = \{1, 2, 3\}. \end{cases} \quad (1)$$

where X is the input vector representing the 11 features of the drift region as shown in table 1, Z_j and H_k are the neurons of hidden layer 1 and hidden layer 2 in which $\{j = 1, 2, \ldots, 32\}$ and $\{k = 1, 2, \ldots, 16\}$, and Y_o is the output vector where $\{o = 1, 2, 3\}$ represents the three locations as shown in Fig. 2. $w_{ji}^{(l)}$ is the weight of the connection between neuron i from layer $l-1$ and neuron j from layer l. Note that the bias items b_j, b_k and b_o are omitted in Fig. 4 for better visibility. Besides, rectified linear unit (ReLU) [28] that ranges from 0 to ∞ is used as an activation function in our DNN structure. Such nonlinear activation provides the model with the capacity needed to model highly sophisticated relations.

For classification tasks, cross-entropy is typically used as the loss function during the training process [29]. Mathematically, this loss function can be expressed as below:

$$L(Y, Y^a) = -\frac{1}{N}\sum_{i=1}^{N}\left(y_i ln(y_i^a) + (1-y_i) ln(1-y_i^a)\right) \quad (2)$$

where y^a is the corresponding set of actual labeled locations, y is the predicted results outputted by the neural network, and N is the number of training samples. Moreover, Adam optimizer, with an adaptive learning rate, is adopted for the gradient descent optimizer for DNN model training [30], [31]. The initial learning rate is set to 0.01, and then it is decreased gradually by a factor of 0.7 after 1000 training iterations so as to make the update small when approaching the optimal solution. Algorithm 1 summarizes the whole procedure for breakdown location prediction using DNN.

Algorithm 1 Algorithm for Breakdown Location Prediction Using DNN

Require: Labeled pair of training data X, Y^a (Y^a denotes to the specific location).

Ensure: Breakdown Location with maximum accuracy.

1: Define $w_{ji}, w_{kj}, w_{ok}, b_j, b_k, b_o$ as the weights and bias items;
2: Define R as the maximum iterations for DNN learning;
3: Intializate $w_{ji}, w_{kj}, w_{ok}, b_j, b_k, b_o$;
4: **for** $t = 1 \rightarrow R$ **do**
5: Calculate the predicted Y according to Eq. (1);
6: Compute the error between Y and golden label Y^a according to Eq. (2);
7: Modify the weights and bias items according to the error;
8: **end for**
9: **return** DNN model

C. DEEP NEURAL NETWORK FOR BV PREDICTION

Fig. 5 shows the DNN architecture for BV prediction of SOI lateral power device. The architecture includes four hidden layers where the details of these layers are shown in Fig. 5. The principle of its implementation is similar to that of the DNN classification model presented in III-B. It establishes a regression model by updating the weights between neurons through information forward propagation and error backpropagation. However, the output of the network should be continuous; hence, a single neuron is used at the output layer containing the predicted BV value. Moreover, unlike the classification model, square loss is used to train the regression model as it is more suitable for this task.

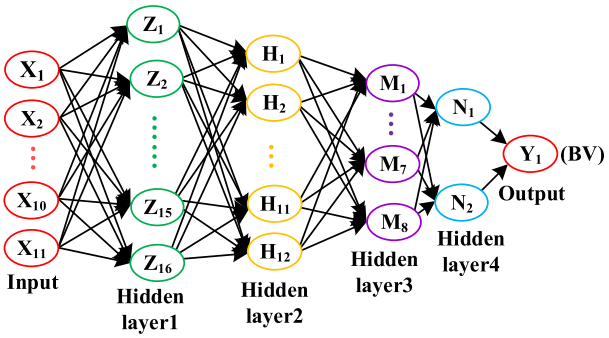


FIGURE 5. Backpropagation DNN architecture for BV prediction.

D. GAUSSIAN PROCESS REGRESSION FOR BV PREDICTION

Most of the machine learning models that are widely adopted fall under the title of parametric model; i.e., the objective is to solve for a set of parameters that define a model which can be used for a regression or a classification task. One example of such models is DNN where the model is defined by the set of weights and bias terms. Despite the fact that these models have achieved successful results in many domains, the parametric nature of these models limits the complexity of the function to be modeled. In fact, with more complicated function, one would resort to a larger set of parameters to better address the task at hand. However, Gaussian process regression falls under the category of non-parametric models with no predefined model template [32]–[34].

At its core, GPR uses the correlation between the test sample and the data in the training set to make the prediction. Hence, the regression model can handle cases with any level of complexity. Moreover, another distinguishing feature for GPR is that it is a probabilistic predictor as opposed to models with point estimates only. In other words, instead of providing a single value as the prediction, GPR provides a distribution as its prediction. This, in fact, can be used to provide a confidence level for the predictions, which has been adopted in other fields like lithography hotspot detection for semiconductor manufacturing [14].

In this section, we briefly review the GPR which consists of model training and prediction in the BV prediction.

1) PREDICTION

Regarding a regression problem $y = f(x) + \xi$ using GPR method, the latent function $f(x)$ is presumed to have a Gaussian process distribution $\mathcal{N}(0, k)$, and ξ is Gaussian noise which has an independent normal distribution with 0 mean and variance σ_N^2. Given a training set with N points, $X = \{x_i, \ldots, x_N\}^T$, where $x_i \in \mathbb{R}^{11}$ is the input vector

with 11 feature, and their corresponding labeled values $y = \{y_i, \ldots, y_N\}^T$, where y_i is the corresponding BV. The set of generating function collection $\{f(x_1), \ldots, f(x_N)\}$ follows a multivariate normal distribution [35]:

$$\{f(x_1), \ldots, f(x_N)\}^T \sim \mathcal{N}(0, K) \tag{3}$$

where K is the $N \times N$ covariance matrix with the (i, j)-th component $K_{ij} = k(x_i, x_j)$ which is used to measure the correlation between structural parameters x_i and x_j.

In order to get the prediction of BV $y_* = \{y_{*1}, \ldots, y_{*M}\}^T$ using the given drift region feature vectors $X_* = \{x_{*1}, \ldots, x_{*M}\}^T$, the joint distribution of our prior over observation y and targets y_* is Gaussian as well and can be described as:

$$P\left(\begin{bmatrix} y \\ y_* \end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} K(X,X) + \sigma_N^2 I_N & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix}\right) \tag{4}$$

where $K(X, X) = K$, $K(X_*, X_*)$ is the $M \times M$ covariance function of testing dataset, and $K(X, X_*)$ and $K(X_*, X)$ are the $N \times M$ and $M \times N$ covariance functions between the input of testing and training structural parameters, respectively.

Thus, according to the discussion above, to calculate the predictive posterior of the testing data, the conditional distribution $p(y_*|y, X, X_*)$ can be expressed as [32]:

$$p(y_*|y, X, X_*) \sim \mathcal{N}(\mu_*, \Sigma_*) \tag{5}$$

The mean and variance of Eq. (5) are shown as follows,

$$\mu_* = K(X_*, X) Ay \tag{6}$$

$$\Sigma_* = K(X_*, X_*) - K(X_*, X) AK(X, X_*) \tag{7}$$

where $A = \left(K(X,X) + \sigma_N^2 I_N\right)^{-1}$, μ_* denotes the mean prediction of BV and Σ_* represents the variance of the predictions. It is interesting to notice that Σ_* indicates the confidence level of the prediction. For each sample, the value of the variance reflects a level of confidence of the model about the prediction.

2) TRAINING

GPR can choose different covariance functions. The covariance function $K(X_*, X)$ denotes the correlation of output at X_* and X that the intuition indicates that the predicted output is more susceptible to the output of nearby inputs in the training data. In our study, the Radial Basis Function (RBF) kernel [36] with variance parameter is employed for GPR, which can be expressed as:

$$k(x_i, x_j) = \sigma^2 exp^{-\frac{\|x_i - x_j\|^2}{2\sigma_*^2}} \tag{8}$$

where σ is the variance parameter which controls the smoothness of the function, σ_* is a bandwidth that controls the radial range of action. The parameter set $\theta = \{\sigma, \sigma_*, \sigma_N\}$ contains the hyperparameters, which are generally obtained by the maximum likelihood function: $argmax_\theta Pr(y/X, \theta)$. By deriving the hyperparameters θ, then using the conjugate gradient method, Newton method, or other optimization methods to minimize the partial derivative, the optimal solution of the hyperparameters can be obtained. Here, the logarithmic likelihood function and its partial derivative form with respect to hyperparameters are as follows:

$$\frac{\partial L}{\partial \theta_i} = \frac{1}{2} trC^{-1} \frac{\partial C}{\partial \theta_i} - \frac{1}{2} y^T C^{-1} \frac{\partial C}{\partial \theta_i} C^{-1} y \tag{9}$$

where $C = K(X, X) + \sigma_N^2 I_N$, $tr(\bullet)$ is the trace of a matrix. Using the optimal solution of those hyperparameters trained by training data, the BV can be obtained according to Eq. (6). Algorithm 2 summarizes the complete procedure for BV prediction using GPR.

Algorithm 2 Algorithm for BV Prediction Using GPR

Require: i^{th} (i = 1,2,3) group training data X,y with specific breakdown location.

Ensure: i^{th} (i = 1,2,3) model with optimal hyperparameters, i.e. σ,σ_*, σ_N.

1: Initialize i^{th} group hyperparameters (σ,σ_*, σ_N);
2: **while** *likelihood* < *threshold* **do**
3: Update σ, σ_* and σ_N using the optimization methods and Eq. (9);
4: **end while**
5: **return** GPR model

IV. EXPERIMENTAL RESULTS

In this section, the efficiency of the proposed two-stage *PowerNet* BV prediction framework is validated with experimental results. First, the training and testing data from TCAD simulator are shown. Second, the data preprocessing technique is presented. Then, the breakdown location prediction accuracy using DNN which comprises the first stage in *PowerNet* is shown, and the accuracy is compared with different classification algorithms. Next, the behavior of the model by examining the training and testing error during the training process is presented. Then, the comparison of the results on BV prediction with different setups, including using single-stage regression or various regression algorithms as the second stage along with the advantage of using GPR are presented. Finally, *PowerNet* is compared to the conventional flow with device simulation in terms of both runtime and the convergence. Our proposed model is implemented in Python on a Linux server with 8-core 3.4 GHz CPU. We use tensorflow library [37] for the DNN model and scikit-learn library [38] for the GPR model.

A. BENCHMARK INFORMATION

The dataset we employed is derived from the commercial device simulator Medici [24]. The feature vector contains 11 structural parameters of the drift region as presented in table 1 and the label contains the BV and the breakdown location. Also, the data for each parameter is sampled from the

TABLE 2. Training and testing dataset from Medici simulator.

Dataset	Total (1510)			
	Train (60%)			Test (40%)
	Class 1	Class 2	Class 3	Class 1-3
	345	496	67	602
Total	908			602

corresponding ranges shown in table 1. A total of 1510 samples is obtained by tool simulation and divided into 6:4, of which 60% is used as a training set and 40% is used as a testing set. We divide the 60% data into three categories in terms of its breakdown location. Note that the breakdown voltages obtained ranged from 20V to 650V. For conventional SOI structures, due to the limitation of vertical voltage withstand, the voltage withstand of SOI high-voltage devices which have really entered the practical stage has not exceeded 600V. Table 2 lists the details of training and testing datasets. The training dataset consists of 3 sets of data labeled with specific breakdown location. The testing dataset is composed of an entire dataset with unknown breakdown location.

B. DATA PREPROCESSING

Different input indicators often have different ranges and dimensional units. For example, the thickness is a few microns, and the concentration can reach 10^{20}. Such situations will affect the results of data analysis. To eliminate the influence of different ranges between indicators, it's imperative to do some data preprocessing before the learning process. Among those data preprocessing methods, the most typical is the data normalization, whose purpose is to limit the data to a specific range, such as [0, 1]. Moreover, this kind of operation brings two advantages, one of which speeds up the gradient descent method to attain the optimal solution, and the other is that normalization can help in improving the prediction accuracy.

C. RESULTS ON BREAKDOWN LOCATION PREDICTION

Fig. 6 shows the number of correctly predicted locations with different algorithms. The blue line is the actual samples count, and the different color blocks show the total number of correctly predicted locations. A simple logistic regression (LR) and support vector classification (SVC) are tried for the breakdown location prediction. Apparently, the red blocks are generally more close to the baseline, which shows that the results of LR and SVC of each location prediction are inferior to that of the DNN methods. Moreover, the use of DNN allows the accuracy of breakdown location prediction to be as high as 97.67%.

D. RESULTS ON BV PREDICTION

1) ERROR FOR MODEL TRAINING PROCESS AND TESTING PROCESS

Fig. 7 shows the BV prediction error of the model training process and corresponding error of test dataset with predicted location of the three location-specific neural networks, respectively. It is important to note that the testing data was only evaluated during the process, and was not

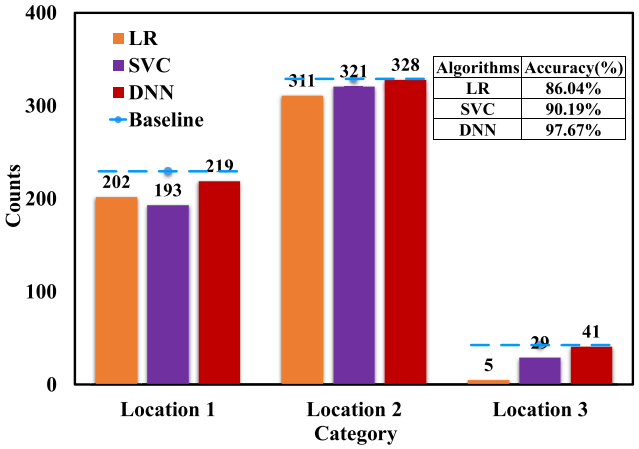


FIGURE 6. The number of correctly predicted breakdown locations using different classification algorithms.

used in any training process. The blue line denotes the BV prediction error during the model training process and the red line denotes the BV prediction error for the test dataset. The error plots validate the convergence of the training process. Besides, both training and testing errors exhibit similar behavior which shows that the model is generalizing well with no overfitting.

2) COMPARISON BETWEEN SINGLE-STAGE REGRESSION AND TWO-STAGE REGRESSION

Table 3 shows the comparison of the results on BV prediction between single-stage regression of our preliminary results in [39] and this method. Regarding single-stage regression, various regression algorithms are adopted to train the BV prediction model, including linear regression (LinReg), support vector regression (SVR), DNN and GPR. Note the prediction results of single-stage regression are obtained using the entire training dataset including samples from training class 1, 2 and 3 to train the model, and then testing is done against the entire testing dataset to obtain the prediction error. For LinReg and SVR, it is evident that the model fails to capture the BV behavior, whereas the predictions from single-stage DNN and GPR are more accurate with error values around 8.8%.

By calculating the total average error according to the data proportion, the final average error for DNN is 3.8% and 3.7% for GPR, which is twice as small as the best result using single-stage regression. Hence, this proves that it is challenging for single model to capture the correlation between BV and structural parameters. The employment of first stage location classification is beneficial for avoiding the sharp change in the dataset, which is more conducive to the breakdown prediction task. The significant improvement of breakdown prediction can be ascribed to the employment of first stage data classification.

3) COMPARISON BETWEEN DIFFERENT REGRESSION ALGORITHMS

Various regression algorithms have been explored for the second stage BV prediction under same first stage location prediction. Fig. 8 shows the experimental results using

TABLE 3. Comparison between single-stage regression and our two-stage method.

Method	Two-stage Regression		Single-stage Regression			
	DNN+DNN	DNN+GPR	SVR	LinReg	DNN	GPR
Prediction Error (%)	3.8	3.7	47.8	41.9	8.6	8.8

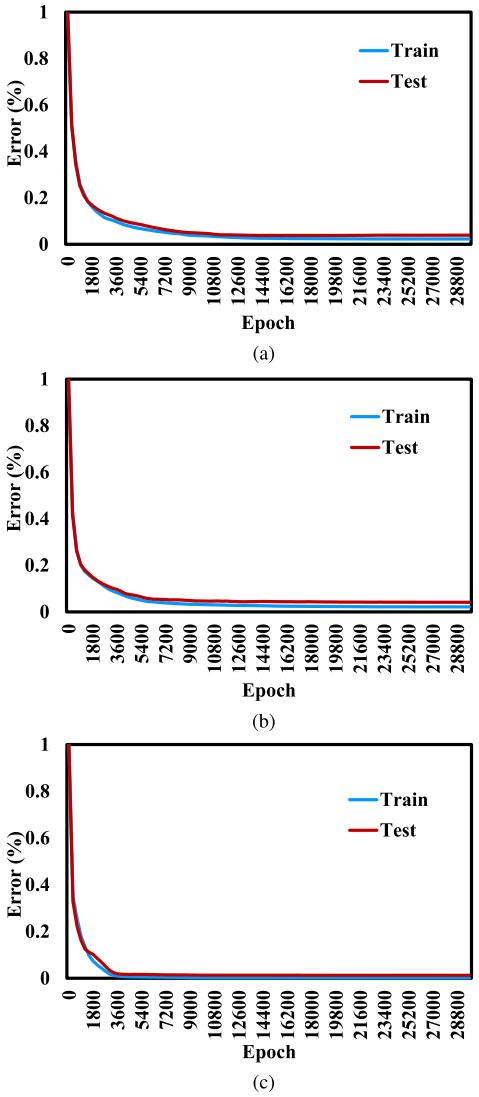


FIGURE 7. BV prediction error during the model training process and testing for the three location-specific neural network, (a) location 1; (b) location 2; (c) location 3.

different algorithms, including LinReg, SVR, DNN, and GPR, which mark different colors. We can observe that the mean prediction error corresponding to LinReg and SVR are all above 10%, especially for class 2, the error values are even larger than 20%. The results show that SVR, with a nonlinear RBF kernel, fails to properly capture the BV behavior. This suggests that the response is highly complex and can not be captured using polynomial or simple nonlinear functions. However, the mean prediction error for each category using DNN or GPR is less than 5%. Thus, we employ DNN or GPR as second stage BV prediction in the *PowerNet* while GPR shows advantages in providing the confidence level of prediction which will be discussed in the following section.

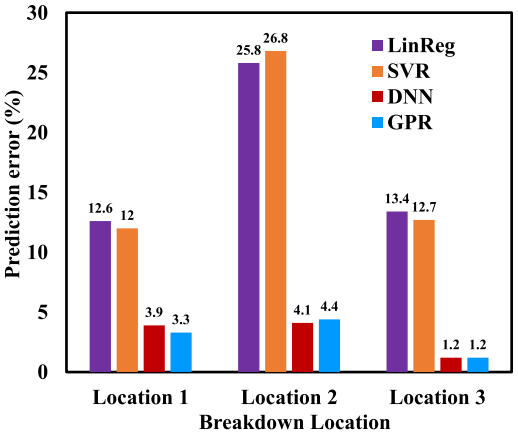


FIGURE 8. Experimental results comparison about BV prediction on the second stage with different regression algorithms.

TABLE 4. Runtime comparison between TCAD simulator and PowerNet.

Runtime	Medici Simulator (s)	PowerNet (s)	
		Breakdown Location	BV
Value	5.6×10^4	0.01	0.03
Total	5.6×10^4	0.04	

4) ADVANTAGES ON USING GPR COMPARED WITH DNN

While the regression accuracies using DNN and GPR are close, GPR can provide a more elaborative prediction than DNN. As discussed in section III-D, the variance obtained from GPR can be used to estimate a confidence interval around the mean prediction. Fig. 9 shows 10 samples of the BV prediction results using GPR as the second stage where the mean prediction along with the 1-*sigma* confidence interval are presented. Shown also in the figure is the golden label (marked in red) which lies within the confidence interval predicted by GPR for all samples.

It is important to note here two important points. The first is that the fact that the label always lies inside the confidence interval predicted by the model reflects the accuracy of the probabilistic predictions provided by the GPR. Secondly, the confidence interval itself can be used by the user as a measure of trust on the prediction mean. In other words, a user will have more trust in a prediction with a tight confidence interval. Hence, GPR can provide more expressive information to the user concerning the prediction.

5) RESULTS ON BV PREDICTION COMPARED WITH TCAD SIMULATOR

Fig. 10 gives an example showing that the BV changes sharply as the doping concentration changes between 11 $\times 10^{15} cm^{-3}$ to 12 $\times 10^{15} cm^{-3}$, which tends to happen in the manufacturing process. Such drastic changes are easily

TABLE 5. Non-convergent simulation under our mesh conditions.

Structural Parameters	t_{ox}	N_d	t_s	L	N_n	P_p	t_n	t_p	r	l_n	l_p
1	0.5	7.50×10^{15}	1.1	77	5.50×10^{20}	5.50×10^{18}	t_s/1.5	t_s	2.4	5.4	6.4
2	2.2	4.90×10^{15}	2.2	69.5	1.48×10^{21}	1.48×10^{19}	t_s/1.75	t_s/1.5	3.9	6.8	8.8
3	3	7.70×10^{15}	3.6	53	1.76×10^{21}	1.76×10^{19}	t_s/1.5	t_s/1.25	2.7	4.6	6.6
...											

TABLE 6. Comparison on the results of tool simulation after mesh adjustment and our model.

Parameter	Tool Simulation by Mesh Adjustment		PowerNet		Prediction Error (%)
	Breakdown Location	Breakdown Voltage	Breakdown Location	Breakdown Voltage	
1	1	90.8	1	92.1	1.4
2	1	258	1	255.7	0.8
3	2	85.3	2	83.5	2.1

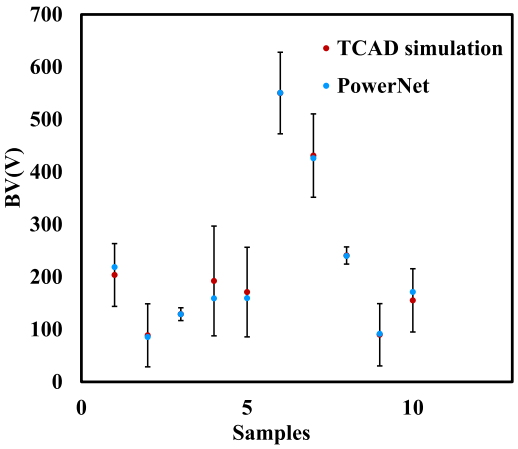


FIGURE 9. Experimental results on 10 samples of BV prediction using GPR as the second stage, where the mean prediction along with the 1-sigma confidence interval are presented.

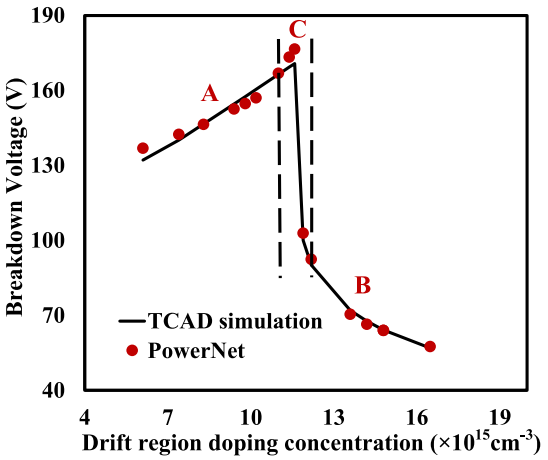


FIGURE 10. Experimental results on BV prediction with various drift region doping concentration using PowerNet compared with TCAD simulator.

accompanied by many parameters, such as t_s, t_{ox}, etc. The red dots are in good agreement with the TCAD simulation, which proves the efficiency of the proposed method. In practice, our proposed approach can provide BV performance estimation for the designer who can make adjustment to limit the effects of variation. For example, the designer would choose design points from regions A and B, which are more robust than points in C. Moreover, these results are given in extremely short period.

E. RUNTIME COMPARISON

Table 4 shows runtime for the total testing data of BV estimation using both the TCAD simulator and *PowerNet*. The TCAD simulation is implemented on a virtual machine with 2-core 2.4 GHz CPU. The inference time of our model prediction is composed of two parts, the time of the breakdown location prediction by the DNN and the time of the BV prediction by the DNN or GPR. Note that the inference time of using DNN or GPR as second stage regression are similar. It is apparent that the BV prediction using machine learning models can dramatically shorten the inference time. The reason why the simulation tool takes a high computational cost to estimate the breakdown performance is that it solves the corresponding physical equation at the set mesh point by gradually increases the applied reverse voltage. The process of steadily increasing the BV and addressing the physical equation is time-consuming, especially when the BV is relatively high. For example, simulating a structure with a BV of 583V takes up to 438s. Instead, breakdown performance can be efficiently predicted through the machine learning framework based on the structural parameters.

F. CONVERGENCE ANALYSIS

Since the simulation tool computes the breakdown performance by solving the physical function at the set mesh point, there are cases where the calculation does not converge. Therefore, the mesh needs to be manually adjusted continuously until convergence is achieved. The problem can be completely solved by our approach. Table 5 shows samples where the simulation does not converge under the condition of our mesh set. In such cases, the mesh needs to be redefined before running a new simulation. Table 6 shows both the results of the device simulation under mesh adjustment and our model using the corresponding structural parameters in table 5. It is evident that our model is robust and can overcome the convergence issue faced by simulation tool. This convergence issue is encountered in our experiments while

using 2D simulation, however, it is much more prevalent in the 3D simulation setup which is associated with an even longer runtime.

V. CONCLUSION

In this paper, a novel breakdown performance prediction framework, *PowerNet*, based on deep neural network, is established. It can efficiently provide breakdown location and BV of a device to evaluate the random variance during the manufacturing process. It utilizes a two-stage modeling technique; the first comprises a breakdown location prediction using DNN; while the second performs the BV prediction based on the classification results using DNN or GPR. The experimental results show that the proposed approach achieves 97.67% breakdown location prediction accuracy and less than 4% average error on the BV prediction. Moreover, the machine learning method significantly reduce the computational cost and is capable of avoiding the convergence problem faced by the simulation tool used for breakdown prediction, which is more convenient for the designers.

Besides, the use of machine learning for SOI lateral power device configurations is a promising direction to move forward. In our future work, we plan to investigate developing machine learning models that are capable of predicting the optimal device structural parameters needed to achieve a given performance target. While this work has introduced machine learning techniques for the prediction of breakdown characteristics in SOI devices, the work can be extended to characterize other devices including bulk silicon, silicon carbide (SiC) and Gallium nitride (GaN). In addition, our proposed framework has great potential to address other electrical performance prediction tasks such as current characteristics, frequency characteristics, noise, and reliability.

REFERENCES

[1] S. J. Mahabadi, S. Rajabi, and J. Loiacono, "A novel partial SOI LDMOSFET with periodic buried oxide for breakdown voltage and self heating effect enhancement," *Superlattices Microstruct.*, vol. 85, pp. 872–879, Sep. 2015.

[2] S. Hu, J. Luo, Y. Jiang, K. Cheng, Y. Chen, J. Jin, J. Wang, J. Zhou, F. Tang, X. Zhou, and P. Gan, "Improving breakdown, conductive, and thermal performances for SOI high voltage LDMOS using a partial compound buried layer," *Solid-State Electron.*, vol. 117, pp. 146–151, Mar. 2016.

[3] E. C. T. Kho, A. D. Hoelke, S. J. Pilkington, D. K. Pal, W. A. W. Z. Abidin, L. Y. Ng, M. Antoniou, and F. Udrea, "200-V lateral superjunction LIGBT on partial SOI," *IEEE Electron Device Lett.*, vol. 33, no. 9, pp. 1291–1293, Sep. 2012.

[4] Y. Chen, P. Gan, Y. Jiang, J. Zhou, X. C. Zhou, S. D. Hu, K. Cheng, and F. Tang, "Improving breakdown performance for novel LDMOS using n+ floating islands in substrate," *Electron. Lett.*, vol. 52, no. 8, pp. 658–659, Apr. 2016.

[5] K. Zeng, A. Vaidya, and U. Singisetti, "1.85 kV breakdown voltage in lateral field-plated Ga_2O_3 MOSFETs," *IEEE Electron Device Lett.*, vol. 39, no. 9, pp. 1385–1388, Sep. 2018.

[6] A. Binder and J.-S. Yuan, "Optimization of an enhancement-mode AlGaN/GaN/AlGaN DHFET towards a high breakdown voltage and low figure of merit," in *Proc. IEEE 5th Workshop Wide Bandgap Power Devices Appl. (WiPDA)*, Oct. 2017, pp. 122–126.

[7] X. Luo, J. Fan, Y. Wang, T. Lei, M. Qiao, B. Zhang, and F. Udrea, "Ultralow specific on-resistance high-voltage SOI lateral MOSFET," *IEEE Electron Device Lett.*, vol. 32, no. 2, pp. 185–187, Feb. 2011.

[8] S. Kumar, E. Goel, K. Singh, B. Singh, M. Kumar, and S. Jit, "A compact 2-D analytical model for electrical characteristics of double-gate tunnel field-effect transistors with a SiO_2/high-k stacked gate-oxide structure," *IEEE Trans. Electron Devices*, vol. 63, no. 8, pp. 3291–3299, Jun. 2016.

[9] J.-F. Yao, Y.-F. Guo, T. Xia, J. Zhang, and H. Lin, "3D analytical model for the SOI LDMOS with alternating silicon and high-k dielectric pillars," *Superlattices Microstruct.*, vol. 96, pp. 95–103, Aug. 2016.

[10] J. Zhang, Y.-F. Guo, Y. Xu, H. Lin, H. Yang, Y. Hong, and J.-F. Yao, "One-dimensional breakdown voltage model of SOI RESURF lateral power device based on lateral linearly graded approximation," *Chin. Phys. B*, vol. 24, no. 2, Feb. 2015, Art. no. 028502.

[11] J. Zhang, Y.-F. Guo, D. Z. Pan, and K.-M. Yang, "A novel 3-D analytical method for curvature effect-induced electric field crowding in SOI lateral power device," *IEEE Trans. Electron Devices*, vol. 63, no. 11, pp. 4359–4365, Nov. 2016.

[12] S. Yuan, B. Duan, Z. Cao, H. Guo, and Y. Yang, "Analytical model of LDMOS with a double step buried oxide layer," *Solid-State Electron.*, vol. 123, pp. 6–14, Sep. 2016.

[13] Y. Lin, M. B. Alawieh, W. Ye, and D. Z. Pan, "Machine learning for yield learning and optimization," in *Proc. IEEE Int. Test Conf. (ITC)*, Oct./Nov. 2018, pp. 1–10.

[14] W. Ye, M. B. Alawieh, M. Li, Y. Lin, and D. Z. Pan, "Litho-GPA: Gaussian process assurance for lithography hotspot detection," in *Proc. Design, Automat. Test Eur. Conf. Exhib. (DATE)*, 2019, pp. 54–59.

[15] H. Yang, S. Li, Y. Ma, B. Yu, and E. F. Young, "GAN-OPC: Mask optimization with lithography-guided generative adversarial nets," in *Proc. 55th ACM/ESDA/IEEE Design Automat. Conf. (DAC)*, Jun. 2018, pp. 1–6.

[16] Z. Xie, Y.-H. Huang, G.-Q. Fang, H. Ren, S.-Y. Fang, Y. Chen, and J. Hu, "RouteNet: Routability prediction for mixed-size designs using convolutional neural network," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design (ICCAD)*, Nov. 2018, pp. 1–8.

[17] W. Ye, M. B. Alawieh, Y. Lin, and D. Z. Pan, "LithoGAN: End-to-end lithography modeling with generative adversarial networks," in *Proc. 56th Annu. Design Automat. Conf.*, 2019, p. 107.

[18] M. B. Alawieh, Y. Lin, Z. Zhang, M. Li, Q. Huang, and D. Z. Pan, "GAN-SRAF: Sub-resolution assist feature generation using conditional generative adversarial networks," in *Proc. 56th Annu. Design Automat. Conf.*, 2019, p. 149.

[19] X. Zeng, Z. Li, W. Gao, M. Ren, J. Zhang, Z. Li, and B. Zhang, "A novel virtual sensing with artificial neural network and k-means clustering for IGBT current measuring," *IEEE Trans. Ind. Electron.*, vol. 65, no. 9, pp. 7343–7352, Sep. 2018.

[20] S. H. Ali, M. Heydarzadeh, S. Dusmez, X. Li, A. S. Kamath, and B. Akin, "Lifetime estimation of discrete IGBT devices based on Gaussian process," *IEEE Trans. Ind. Appl.*, vol. 54, no. 1, pp. 395–403, Jan. 2018.

[21] H. Carrillo-Nunez, N. Dimitrova, A. Asenov, and V. Georgiev, "Machine learning approach for predicting the effect of statistical variability in Si junctionless nanowire transistors," *IEEE Electron Device Lett.*, vol. 40, no. 9, pp. 1366–1369, Sep. 2019.

[22] X. Luo, B. Zhang, and Z. Li, "A new structure and its analytical model for the electric field and breakdown voltage of SOI high voltage device with variable-k dielectric buried layer," *Solid-State Electron.*, vol. 51, no. 3, pp. 493–499, Mar. 2007.

[23] W. Yang, X. Cheng, Y. Yu, Z. Song, and D. Shen, "A novel analytical model for the breakdown voltage of thin-film SOI power MOSFETs," *Solid-State Electron.*, vol. 49, no. 1, pp. 43–48, Jan. 2005.

[24] Technology Modeling Associates. (1996). *Medici: Two-dimensional Device Simulation Program, Version 2.2: User's Manual. Technology Modeling Associates, Incorporated*. [Online]. Available: https://books.google.com/books?id=_jknywAACAAJ

[25] R. Rojas-Moraleda, N. A. Valous, A. Gowen, C. Esquerre, S. Härtel, L. Salinas, and C. O'Donnell, "A frame-based ANN for classification of hyperspectral images: Assessment of mechanical damage in mushrooms," *Neural Comput. Appl.*, vol. 28, no. S1, pp. 969–981, Dec. 2017.

[26] N.-D. Hoang and Q.-L. Nguyen, "A novel method for asphalt pavement crack classification based on image processing and machine learning," *Eng. Comput.*, vol. 35, no. 2, pp. 487–498, Apr. 2019.

[27] A. Gautam, V. Bhateja, A. Tiwari, and S. C. Satapathy, "An improved mammogram classification approach using back propagation neural network," in *Data Engineering and Intelligent Computing*. Singapore: Springer, 2018, pp. 369–376.

[28] A. F. Agarap, "Deep learning using rectified linear units (ReLU)," 2018, *arXiv:1803.08375*. [Online]. Available: https://arxiv.org/abs/1803.08375

[29] Z. Zhang and M. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 8778–8788.
[30] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980
[31] M. D. Zeiler, "ADADELTA: An adaptive learning rate method," 2012, *arXiv:1212.5701*. [Online]. Available: https://arxiv.org/abs/1212.5701
[32] C. E. Rasmussen, "Gaussian processes in machine learning," in *Summer School on Machine Learning*. Berlin, Germany: Springer, 2003, pp. 63–71.
[33] C. K. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning*, vol. 2, no. 3. Cambridge, MA, USA: MIT Press, 2006.
[34] E. V. Bonilla, K. M. Chai, and C. Williams, "Multi-task Gaussian process prediction," in *Proc. Adv. Neural Inf. Process. Syst.*, 2008, pp. 153–160.
[35] M. M. Sawant and K. Bhurchandi, "Hierarchical facial age estimation using Gaussian process regression," *IEEE Access*, vol. 7, pp. 9142–9152, 2019.
[36] S. Sun, G. Zhang, C. Wang, W. Zeng, J. Li, and R. Grosse, "Differentiable compositional kernel learning for Gaussian processes," 2018, *arXiv:1806.04326*. [Online]. Available: https://arxiv.org/abs/1806.04326
[37] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, and M. Isard, "TensorFlow: A system for large-scale machine learning," in *Proc. OSDI*, vol. 16, 2016, pp. 265–283.
[38] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, and V. Dubourg, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.
[39] J. Chen, Y. Guo, Y. Lin, M. B. Alawieh, M. Zhang, J. Zhang, and D. Z. Pan, "Breakdown voltage prediction of SOI lateral power device using deep neural network," in *Proc. Cross Strait Quad-Regional Radio Sci. Wireless Technol. Conf. (CSQRWC)*, Jul. 2019, pp. 1–3.

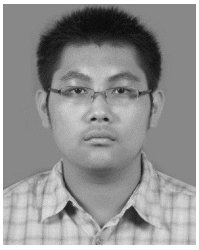
JING CHEN (Student Member, IEEE) received the B.S. degree in microelectronics from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2016, where she is currently pursuing the Ph.D. degree. She is also a Visiting Scholar with The University of Texas at Austin. Her research interests include machine learning, design for manufacturability, and power device.

MOHAMED BAKER ALAWIEH (Student Member, IEEE) received the B.S. degree in electrical and computer engineering from the American University of Beirut, Beirut, Lebanon, in 2014, and the M.S. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, USA, in 2016. He is currently pursuing the Ph.D. degree in electrical and computer engineering with The University of Texas at Austin, Austin, TX, USA. His current research interests include machine learning and computer-aided design for VLSI.

YIBO LIN (Member, IEEE) received the B.S. degree in microelectronics from Shanghai Jiaotong University, Shanghai, China, in 2013, and the Ph.D. degree in electrical and computer engineering from The University of Texas at Austin, in 2018.

He is currently an Assistant Professor with the Department of Computer Science, Peking University associated with the Center for Energy-Efficient Computing and Applications (CECA), Beijing, China. His research interests include physical design, machine learning applications, emerging technology in VLSI CAD, and hardware security. He was a recipient of the Best Paper Awards at DAC 2019, Integration, the VLSI Journal 2018, and SPIE Advanced Lithography Conference 2016.

MAOLIN ZHANG (Student Member, IEEE) received the B.S. degree in microelectronics from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 2016. He is currently pursuing the Ph.D. degree with the Nanjing University of Posts and Telecommunications, Nanjing, China. His research interests include nano-scale device design and modeling.



JUN ZHANG received the Ph.D. degree in microelectronics from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2018. He is currently a Postdoctoral Scholar with the College of Electronic and Optical Engineering, Nanjing University of Posts and Telecommunications. His current research interests include semiconductor power devices, microelectronics devices reliability, and RF and power integrated circuits and systems.

YUFENG GUO (Member, IEEE) received the Ph.D. degree in microelectronics and solid-state electronics from the University of Electronic Science and Technology of China, Chengdu, China, in 2005. He is currently a Vice Director of the National and Local Joint Engineering Laboratory of RF Integration and Micro-Assembly Technology, NJUPT, Nanjing, China. He has published more than 180 articles in refereed journals and conferences and holding over 20 Chinese patents. His current research interests include semiconductor power devices, microelectronics devices reliability, and RF and power integrated circuits and systems.

DAVID Z. PAN (Fellow, IEEE) received the B.S. degree from Peking University, and the M.S. and Ph.D. degrees from the University of California, Los Angeles (UCLA).

From 2000 to 2003, he was a Research Staff Member with the IBM T. J. Watson Research Center. He is currently an Engineering Foundation Professor with the Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX, USA. He has published over 360 journal articles and refereed conference papers, and is the holder of eight U.S. patents. His research interests include cross-layer nanometer IC design for manufacturability, reliability, security, machine learning and hardware acceleration, design/CAD for analog/mixed signal designs, and emerging technologies. He has received a number of prestigious awards for his research contributions, including the SRC Technical Excellence Award in 2013, DAC Top 10 Author in Fifth Decade, DAC Prolific Author Award, ASP-DAC Frequently Cited Author Award, 18 Best Paper Awards at premier venues (ASPDAC 2020, DAC 2019, GLSVLSI 2018, VLSI Integration 2018, HOST 2017, SPIE 2016, ISPD 2014, ICCAD 2013, ASPDAC 2012, ISPD 2011, IBM Research 2010 Pat Goldberg Memorial Best Paper Award, ASPDAC 2010, DATE 2009, ICICDT 2009, SRC Techcon in 1998, 2007, 2012, and 2015) and 15 additional Best Paper Award finalists, Communications of the ACM Research Highlights, in 2014, UT Austin RAISE Faculty Excellence Award, in 2014, and many international CAD contest awards, among others.