Solid-State Electronics 199 (2023) 108505

Contents lists available at ScienceDirect

Solid State Electronics

journal homepage: www.elsevier.com/locate/sse

Surrogate models for device design using sample-efficient Deep Learning ☆

Rutu Patel *, Nihar R. Mohapatra, Ravi S. Hegde

Indian Institute of Technology Gandhinagar, Gandhinagar, 382355, Gujarat, India

ARTICLE INFO

Keywords:
Deep Neural Networks
Active learning
LDMOSFET
Off-state breakdown voltage
Specific on resistance

ABSTRACT

Generation of training dataset for machine learning-based device design algorithm is expensive. To address this, we propose an active learning approach. Its efficiency is demonstrated through a Deep Neural Network (DNN) based Laterally Diffused Metal Oxide Semiconductor Field-effect Transistor (LDMOSFET) off-state breakdown voltage ($BV_{DS,off}$) and specific on-resistance (R_{sp}) predictor. Our results show the possibility of ~50% reduction in the training dataset size without compromising the baseline accuracy. Specifically, we compared eight sampling techniques and found that Informative-Query by Committee (I-QBC) and Diverse Informative-Greedy Sampling (DI-GS) techniques work best with ~ 1.87% Euclidean Norm of Prediction Error (ENPE).

1. Introduction

Technology Computer Aided Design (TCAD) tools, which solve the physical equations at set mesh points, have been developed and used over the years to reduce semiconductor device design time and cost. However, the computational time for design of the devices with a large number of mesh points is high (e.g. for high electric field simulations to predict $BV_{DS,off}$ of LDMOSFETs). To tackle such bottlenecks, data-driven surrogate models which mimic TCAD tools are being developed [1–6]. Popularly, DNNs trained using supervised learning (labeled samples from the input feature space) are used. The prediction accuracy of these models improve as the dataset size increases. However, generating a large dataset is computationally expensive and not always feasible. Thus, developing surrogate models for complex devices with a large number of input features is a challenge.

Active learning [7] – a technique of choosing the best samples from a pool such that the accuracy of the surrogate model improves – was proposed by Dongrui Wu [8]. It has been recently used for the inverse design of photonic nanostructures [9]. Three criteria were suggested to choose/label the efficient samples from a pool a finite pool- Informativeness (I), Representativeness (R) and Diversity (D). In this work, eight different sampling techniques which promise to reduce the training dataset size are evaluated: I-GS (Informative-Greedy Sampling), I-QBC (Informative-Query by Committee), DI-GS (Diverse Informative-Greedy Sampling), DI-QBC (Diverse Informative-Query by Committee), R-GS (Representative-Greedy Sampling), R-QBC (Representative-Query by Committee), DR-GS (Diverse Representative-Greedy Sampling) and DR-QBC (Diverse Representative-Query by Committee). Using these techniques, a DNN based LDMOSFET $BV_{DS,off}$ and R_{sp} predictor (surrogate model) is developed and for the first time a sample-efficient algorithm is proposed for semiconductor device design. The experiments performed on training datasets with different sizes reveal that up to ~50% reduction in training dataset size can be achieved without affecting the baseline ENPE (accuracy achieved without using active learning based sampling techniques). The I-QBC and DI-GS techniques outperform the others. Such a reduction in training dataset size promises to open new frontiers for inverse design of complex semiconductor devices.

2. Computational framework

Fig. 1 shows the LDMOSFET structure considered in this work. The training and the test datasets are generated by simulating this structure in sprocess and sdevice tools of the Sentaurus TCAD suite [10]. The hydrodynamic carrier transport model is used for simulating the electrical characteristics. The Van Overstraeten–De Man model is used for electron and hole impact ionization (II) and breakdown voltage simulations. The Shockley–Read–Hall and Auger recombination models are included to account for carrier generation and recombination. The band gap narrowing model for silicon, doping dependent Masetti mobility model, Lombardi surface mobility degradation model at silicon-oxide interfaces and high-field mobility saturation models are also used for accurate device simulation. With prior experience of physics based design approach [11], seven design parameters (mentioned in Table 1) are chosen to form the input feature space. Their ranges are varied with respective resolutions so that the two output variables, $BV_{DS,off}$

☆ The review of this paper was arranged by Francisco Gamiz.

* Corresponding author.
E-mail address: patel.rutu@iitgn.ac.in (R. Patel).

https://doi.org/10.1016/j.sse.2022.108505

Available online 5 November 2022

Table 1
Design parameters, their ranges and resolutions.

Design parameter	Range	Resolution
Implant A Dose (D_1)	0.8–1.2 × 10^{12} cm^{-2}	0.1 × 10^{12} cm^{-2}
Implant A Energy (E_1)	160–220 keV	20 keV
Implant B Dose (D_2)	0.8–1.2 × 10^{12} cm^{-2}	0.1 × 10^{12} cm^{-2}
Implant B Energy (E_2)	260–320 keV	20 keV
Drift length (L_D)	1.5–9 μm	0.2 μm
Field plate length (L_{FP})	0.5–4 μm	0.2 μ
Field plate dielectric thickness (t_{FP})	0.5–0.8 μm	0.1 μm
Design constraints: E_1<E_2, L_{FP}<L_D/2		

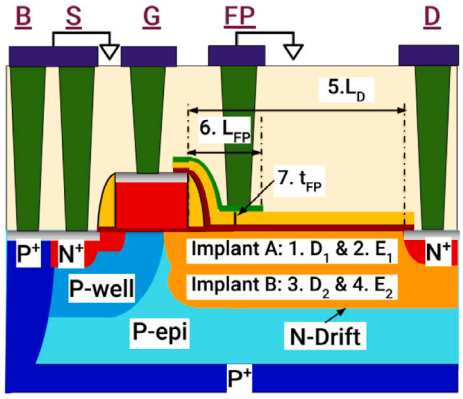

Fig. 1. Schematic of the LDMOSFET structure with the device design parameters (described in Table 1). These parameters affect $BV_{DS,off}$ and R_{sp}.

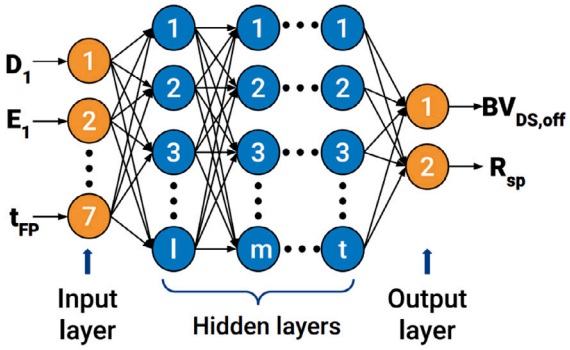

Fig. 2. The DNN based predictor (surrogate model) with seven nodes in the input layer (device design parameters in Table 1) and two nodes in the output layer ($BV_{DS,off}$,R_{sp}).

& R_{sp}, vary between 40–160 V and 90–430 mΩ mm^2 respectively and the generated samples have R_{sp}-$BV_{DS,off}$ tradeoff close to the theoretical Silicon limit (Fig. 8). These input and output parameters combine to form the DNN based surrogate model shown in Fig. 2. Also, they are scaled between 0 and 1 before training the DNN. The scaling strategy is given by, $x_{scaled} = (x_{original} - x_{min})/(x_{max} - x_{min})$

Here, x_{min} and x_{max} are the minimum and maximum values of the parameters. Rectified linear unit (ReLU) is used as the activation function. Mean Absolute Relative Error (MARE), is used as the loss function and it is defined as,

$$MARE = L(Y, Y^a) = \frac{1}{N}\sum_{i=0}^{N}\left|\frac{y_i^a - y_i}{y_i^a}\right|$$

Here, y_i^a is the actual value and y_i is the DNN predicted value of the ith sample in a dataset of N training samples. Adam optimizer, with an adaptive learning rate, is used for DNN model training. ENPE plotted for separately generated test dataset of 300 Latin Hypercube Samples (LHS), is chosen as the accuracy figure-of-merit (FoM) to compare different sampling techniques and structures of DNN predictor. It is defined as,

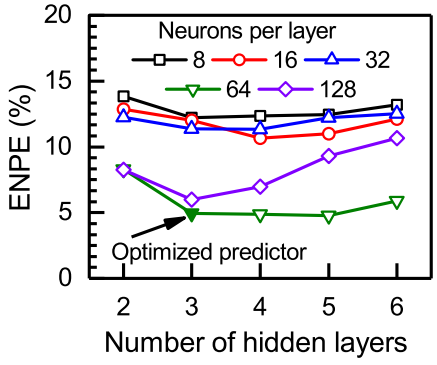

Fig. 3. ENPE vs. number of hidden layers with different numbers of neurons per layer plot. DNN with 3 hidden layers and 64 neurons per layer which shows the lowest ENPE is fixed as the optimized predictor.

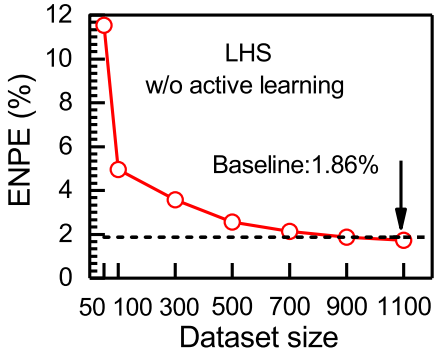

Fig. 4. ENPE vs. dataset size plot shows that ENPE reduces by increasing training dataset size, for the chosen optimized predictor. The baseline ENPE is 1.86% for 1100 LHS samples w/o active learning.

$$ENPE(\%) = \sqrt{PE^2_{BV_{DS,off}} + PE^2_{R_{sp}}}$$

$$PE(\%) = \frac{100}{300}\sum_{i=0}^{300}\left|\frac{y_i^a - y_i}{y_i^a}\right|$$

The configuration of DNN based predictor, for ENPE comparison of different sampling techniques, is first optimized such that a balance between under fitting and over fitting is achieved. Appropriate number of hidden layers and neurons per layer are chosen by performing experiments on a small training dataset of 100 LHS. As evident from Fig. 3, the optimized predictor which provides the minimum ENPE has 3 hidden layers with 64 neurons each. Using this optimized predictor, experiments with increased training dataset size are performed. As shown in Fig. 4, it is observed that the ENPE reduces by increasing the training dataset size and ultimately saturates to ~ 1.86% for 1100 LHS. This is considered as the baseline ENPE (without using active learning based sampling techniques).

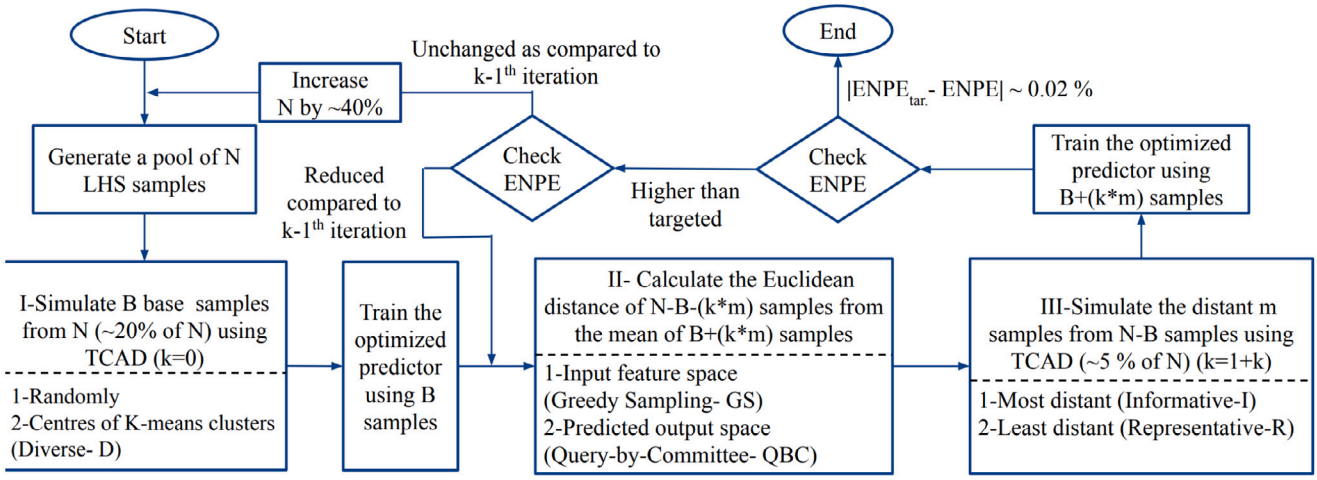

Fig. 5. Flow chart depicting the efficient sampling algorithm. 8 different sampling techniques are proposed, considering the possible methods of (I) choosing B, (II) calculating Euclidean distance and (III) choosing the distant m samples.

3. Sample-efficient algorithm

The sample-efficient algorithm depicted in the flowchart (Fig. 5), is developed to reduce the dataset size without compromising the baseline ENPE. As suggested in [8], first a pool of *N* LHS samples is generated. Next, depending on the possible two methods: (1) randomly or (2) using k-means clustering, B base samples (~20% of N) are selected (I) from the pool and are simulated in TCAD. The second method adds diversity to the base samples as it forms B clusters of N/B samples each and then selects samples which are nearest to their centroids. The optimized predictor is then trained using these B labeled samples. Further, the samples are iteratively added from the remaining unlabeled samples of the pool.

Euclidean distance of the remaining unlabeled (N-B-k*m) samples is calculated from the mean of labeled (B+k*m) samples. Here, k is the number of current iterations and m is the number of labeled samples added per iteration (~5% of N). It is calculated in one of the two spaces (II) and so it leads to two methods: (1) Greedy sampling (GS) in input feature space and (2) Query-by-Committee (QBC) in predicted output space. Once the sampling space is selected, one chooses to add the labeled samples based on their quality of (1) informativeness or (2) diversity i.e. by selecting the most distant or the least distant samples from the mean respectively (III). m additional samples are simulated in TCAD, every iteration, and the optimized predictor is then trained using the total B+(k*m) samples. If the targeted ENPE is achieved ($|ENPE_{tar.}$-ENPE| ~ 0.02%), the developed training dataset is finalized and the algorithm terminates. If not, ENPE is further checked. If the ENPE does not saturate, the process of adding more samples from the unlabeled samples is continued by increasing the number of iterations. If the ENPE is saturated, a pool with higher *N* (~40% more than the previous N) is selected and all the steps are repeated. Note that increasing B instead of *N* is not an option as it will label majority of the samples from the pool without judging their informativeness and representativeness. Further, with higher N, B and m also increase so that targeted ENPE is achieved with minimum iterations.

Combining different methods for steps (I), (II) and (III) of the sample-efficient algorithm, we propose eight techniques which promise to reduce training dataset size without affecting the baseline ENPE. These are as elaborated below,

A. Informative-Greedy Sampling (I-GS): Base B samples are randomly selected and labeled from a pool of *N* LHS samples. m most distant samples in the input feature space (farthest from the mean of previously labeled samples) are added per iteration from the unlabeled samples.
B. Informative-Query by Committee (I-QBC): Same steps are followed as in I-GS, but the predicted output space is chosen as the sampling space.
C. Diverse-Informative Greedy Sampling (DI-GS): Base B samples are selected as the samples which are nearest to the centroids of B k-means clusters. It adds diversity to the selected samples. The additional samples are selected as the most distant unlabeled samples from the mean of previously labeled samples, in the input feature space.
D. Diverse Informative-Query by Committee (DI-QBC): Same steps are followed as in DI-GS, but the predicted output space is chosen as the sampling space.
E. Representative-Greedy Sampling (R-GS): Base B samples are randomly chosen from the pool of *N* LHS samples. Further, the least distant or the nearest unlabeled samples from the mean of previously labeled samples, in input feature space, are added in every iteration.
F. Representative-Query by Committee (R-QBC): Same steps as in R-GS technique are followed but the predicted output space is chosen as the sampling space.
G. Diverse Representative Greedy Sampling (DR-GS): B samples which are nearest to the centroids of B k-means clusters are selected. Least distant m samples in input feature space are added per iteration.
H. Diverse Representative-Query by Committee (DR-QBC): Same steps as in DR-GS technique are followed but the predicted output space is chosen as the sampling space.

Selecting each of the above sample-efficient techniques, experiments are performed to generate training dataset of minimum size such that the targeted baseline ENPE is achieved. Results are discussed in the next section.

4. Results and discussion

The ENPE for all the eight techniques is extracted for the test dataset of size 300. First, by performing the experiments on an LHS pool of size N=500 (Fig. 6a & d). B=100 and m=25 were chosen so that size of training dataset is 200 at the end of 4 iterations. ENPE reduces Due to the increase in dataset size, ENPE for all the 8 techniques reduces. However, the ENPE saturates to the lowest value of ~ 3.12% for the I-QBC technique. To further reduce the ENPE, second dataset with N=700, B=150 and m=50 is considered (Fig. 6b & e). After 4 iterations, the training dataset of size 350 could achieve the lowest ENPE of ~ 2.44% using DI-QBC sampling technique but it still remains

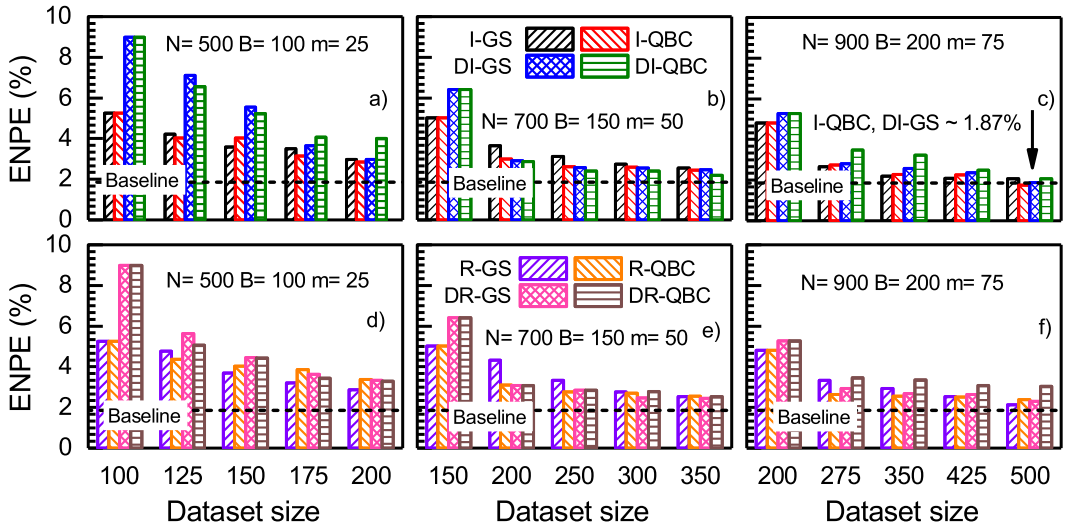

Fig. 6. ENPE of the predictors trained using 8 active learning based sampling techniques for 3 different datasets. It is calculated for the test dataset of size 300. N, B and m increase from (a) to (c) and from (d) to (e). I-QBC and DI-GS sampling techniques achieve ~ 1.87% ENPE with total 500 samples (N=900, B=200 and m=75) as compared to 1100 LHS samples w/o active learning.

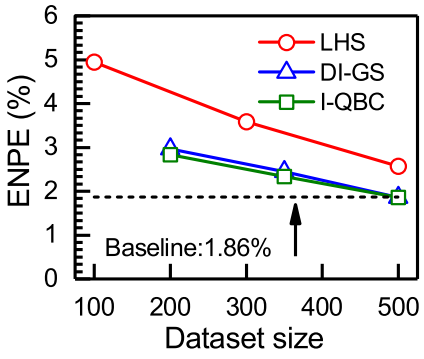

Fig. 7. ENPE vs dataset size plot shows that the LHS under performs when compared with the most efficient DI-GS and I-QBC sampling techniques. The difference between the ENPEs is more evident for smaller training dataset sizes.

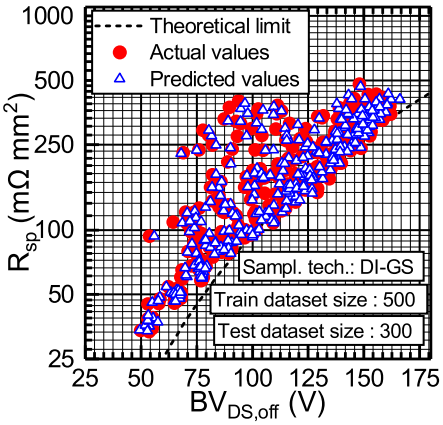

Fig. 8. The R_{sp} vs. $BV_{DS,off}$ plot of actual and predicted values of the test dataset. The most efficient DI-GS sampling technique with total 500 samples is used to train the predictor.

higher than the targeted baseline ENPE. The third dataset with N=900, B=200 and m=75 (Fig. 6c & f) is found to be sufficient to achieve baseline ENPE of ~ 1.87%. Specifically, the I-QBC and DI-GS techniques work best. ENPE achieved with 500 training samples labeled using sample-efficient algorithm is same as that of 1100 LHS without active learning. The training dataset size is reduced by ~50% and thus the computational cost of dataset generation is halved. Fig. 7 compares ENPEs of the most efficient DI-GS and I-QBC sampling techniques with that of LHS. It clearly shows that the proposed techniques perform better than LHS, especially for smaller training dataset sizes.

The surrogate model is further trained with the 500 samples labeled using the DI-GS technique for N=900, B=200 and m=75. For the test dataset of size 300, Fig. 8 shows the R_{sp}-$BV_{DS,off}$ trade-off. As the ENPE is ~ 1.87%, the actual and predicted values are in good agreement.

5. Conclusion

To summarize, a sample-efficient algorithm for device design surrogate model is developed using eight different techniques. It is demonstrated by predicting $BV_{DS,off}$ and R_{sp} of LDMOSFETs. 50% reduction in the training dataset size is achieved by the I-QBC and DI-GS efficient sampling techniques without compromising the baseline ENPE of ~ 1.86%. These benefits can be leveraged by using surrogate models in the inverse design of devices with large number of design parameters.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Nihar R. Mohapatra reports financial support was provided by Science and Engineering Research Board, SERB. Nihar R. Mohapatra reports a relationship with Science and Engineering Research Board, SERB that includes: funding grants.

Data availability

Data will be made available on request.

Acknowledgment

This work is supported by funding from Science and Engineering Research Board, SERB (Department of Science and Technology, Government of India) through their MATRICS scheme.

References

[1] Chen J, Guo Y, Alawieh MB, Zhang M, Zhang J, Pan DZ. An Efficient Automatic Structure Design Method of Silicon-on-Insulator Lateral Power Device Considering RESURF Constraint. IEEE Trans Electron Devices 2021;68(9):4593–7. http://dx.doi.org/10.1109/TED.2021.3101181.

[2] Mehta K, Raju SS, Xiao M, Wang B, Zhang Y, Wong HY. Improvement of TCAD Augmented Machine Learning Using Autoencoder for Semiconductor Variation Identification and Inverse Design. IEEE Access 2020;8:143519–29. http://dx.doi.org/10.1109/ACCESS.2020.3014470.

[3] Gangi H, Taguchi Y, Nakata K, Nemoto H, Kobayashi Y, Inokuchi T, et al. Design Optimization of Multiple Stepped Oxide Field Plate Trench MOSFETs with Machine Learning for Ultralow On-resistance. In: 2021 33rd international symposium on power semiconductor devices and ICs. 2021, p. 151–4. http://dx.doi.org/10.23919/ISPSD50666.2021.9452194.

[4] Myung S, Kim J, Jeon Y, Jang W, Huh I, Kim J, et al. Real-time TCAD: a new paradigm for TCAD in the artificial intelligence era. In: 2020 international conference on simulation of semiconductor processes and devices. 2020, p. 347–50. http://dx.doi.org/10.23919/SISPAD49475.2020.9241622.

[5] Chen J, Alawieh MB, Lin Y, Zhang M, Zhang J, Guo Y, et al. Powernet: SOI lateral power device breakdown prediction with deep neural networks. IEEE Access 2020;8:25372–82. http://dx.doi.org/10.1109/ACCESS.2020.2970966.

[6] Myung S, Jang W, Jin S, Choe JM, Jeong C, Kim DS. Restructuring TCAD system: teaching traditional TCAD new tricks. In: 2021 IEEE international electron devices meeting. 2021, p. 18.2.1–4. http://dx.doi.org/10.1109/IEDM19574.2021.9720616.

[7] Settles B. Active learning literature survey. University of Wisconsin-Madison Department of Computer Sciences; 2009.

[8] Wu D. Pool-Based Sequential Active Learning for Regression. IEEE Trans Neural Netw Learn Syst 2019;30(5):1348–59. http://dx.doi.org/10.1109/TNNLS.2018.2868649.

[9] Hegde R. Sample-efficient deep learning for accelerating photonic inverse design. OSA Continuum 2021;4(3):1019–33. http://dx.doi.org/10.1364/OSAC.420977.

[10] Sentaurus TCAD User Manual Version P -2019.03, Mountain View, CA, USA. 2019.

[11] Patel R, Mohapatra NR. Novel Step Field Plate RF LDMOS Transistor for Improved BV_{DS} - R_{ON} Tradeoff and RF Performance. IEEE Trans Electron Devices 2022;1–7. http://dx.doi.org/10.1109/TED.2022.3182296.