TCAD-Augmented Machine Learning With and Without Domain Expertise

Harsaroop Dhillon, Kashyap Mehta, Ming Xiao, Boyan Wang, Yuhao Zhang, *Member, IEEE*, and Hiu Yung Wong, *Senior Member, IEEE*

***Abstract*— In this article, using experimental data, we demonstrate that the technology computer-aided design (TCAD) is a very cost-effective tool to generate the data to build machine learning (ML) models for semiconductor device and process characterization. Characterization of the emerging ultra wide bandgap gallium oxide (Ga_2O_3) Schottky barrier diode (SBD) is used as an example. Machines are trained by using only TCAD *I–V*'s and then used to deduce the effective Schottky metal work function (WF) and ambient temperature (*T*) of an experimental SBD based on its *I–V*. Besides noise, the experimental device also suffers from relatively large variations in drift layer thickness and doping concentrations. Both ML models with domain expertise (WDE) and without domain expertise (WoDE) are studied and compared. The ML model WDE requires the use of device knowledge to extract relevant features (e.g., subthreshold slope and turn-on voltage) for ML. The ML model WoDE obviates such a requirement and can be extended to cases where domain expertise is difficult to apply. Denoising autoencoder is used in the WoDE case. We showed that with only 500 TCAD *I–V*'s, we can train machines WDE and WoDE that can deduce the experimental device WF and *T* reasonably well. In particular, the ML model WoDE has an acceptable prediction accuracy despite the noise and additional variations in the experimental device.**

***Index Terms*— Autoencoder (AE), gallium oxide, machine learning (ML), simulation, technology computer-aided design (TCAD), ultra wide bandgap.**

I. Introduction

As COMPUTERS become more powerful and low-cost, it is more and more feasible to use technology computer-aided design (TCAD) to generate data to train and build machines for various purposes. For example, ~1 50 000 TCAD current–voltage (I–V) curves can be generated using one server in two weeks [1]. This is called TCAD-augmented [2]–[4] or TCAD-enabled [5] machine learning (ML) because the TCAD simulation data are used to train a machine when the experimental data are scarce. Usually, TCAD simulation data (e.g., I–V) are generated under a well-controlled environment with varying parameters (e.g., effective metal work function (WF) [1]–[6], epitaxial layer thickness [2], defect location [5], gate lengths [6], etc.). A machine is then trained using the TCAD data to learn the correlation between the parameters and the I–V curves. For some given experimental/simulation I–V curves, the machines can then deduce the defect properties [5], reverse engineer the structural information [2], [3], or perform the process variation analysis for the corresponding semiconductor devices or fabrication processes [1].

There are two ways to perform ML using TCAD data, namely with and without domain expertise. In the with domain expertise (WDE) case, device physics knowledge is used to extract the relevant features from the I–V curves, such as the threshold voltage (V_{TH}), subthreshold slope (SS), critical dimensions, etc., [5]–[8], for ML. In the without domain expertise (WoDE) case, the I–V curves are used as the input features directly for ML [1], [4]–[6]. The ML model WoDE avoids the use of device physics and lets the machine discover the physics in the hyperdimensional I–V space itself [6]. This not only makes the process easier by obviating the domain expertise but also is necessary in some cases when domain expertise is not available. However, the machine trained by TCAD data in the WoDE case usually suffers from over-fitting issues, in which the machine produces spurious results when it is applied to experimental data (e.g., negative thickness and negative temperature [1]–[4]). We showed that it is possible to use principal component analysis (PCA) [1], noise [3], or autoencoder (AE) [4] to alleviate the problem.

However, hitherto, there is no study to compare the TCAD-augmented ML models WDE and WoDE that applied to real experimental data. Recently, we present the first work, to the best of our knowledge, that applies the TCAD-augmented ML to real experimental data [1]. Despite the initial demonstration, a few issues remained in the ML models and performance: 1) ML model WDE produces spurious results when it is applied to experimental data; 2) PCA was used in the ML model WoDE but led to unsatisfactory ML performance (temperature prediction range is ~40 K for a given 110 K range data and the values have offset); and 3) large amount of TCAD data (~1 50 000) was used to train the model, which is relatively costly in computation and time.

Manuscript received March 3, 2021; revised April 11, 2021; accepted April 12, 2021. Date of publication April 29, 2021; date of current version October 22, 2021. The review of this article was arranged by Editor F. Bonani. *(Harsaroop Dhillon and Kashyap Mehta contributed equally to this work.) (Corresponding author: Hiu Yung Wong.)*

Harsaroop Dhillon, Kashyap Mehta, and Hiu Yung Wong are with the Department of Electrical Engineering, San Jose State University, San Jose, CA 95192 USA (e-mail: hiuyung.wong@ieee.org).

Ming Xiao, Boyan Wang, and Yuhao Zhang are with the Center for Power Electronics Systems, The Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061 USA.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TED.2021.3073378.

Digital Object Identifier 10.1109/TED.2021.3073378

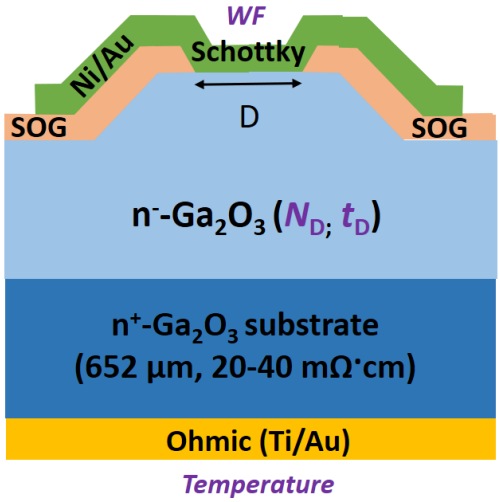

Fig. 1. Schematic of the fabricated vertical Ga_2O_3 power SBDs.

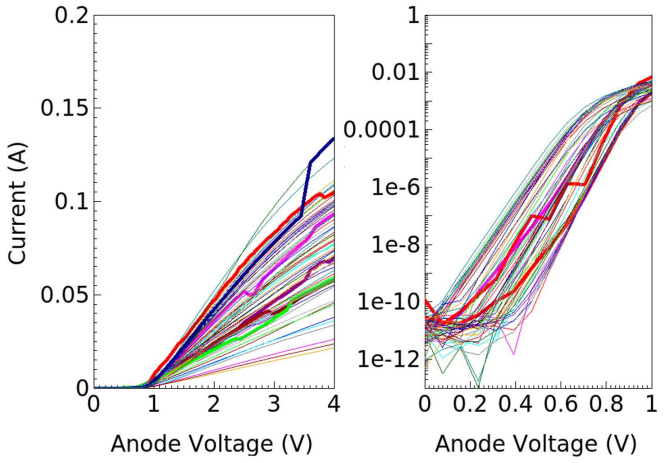

Fig. 2. Collection of 66 experimental *I*–*V* curves measured on 20+ devices at three different temperatures. Left: linear scale. Right: logarithmic scale. The *I*–*V* curves with large noise or abnormalities are highlighted, which are included in the prediction in ML.

In this article, for the first time, we compared the performance of the ML models WDE and WoDE by using the experimental data of the fabricated Ga_2O_3 Schottky barrier diodes (SBDs) in [1]. We show that by using the maximum domain expertise to extract the relevant features, the trained machine performs well with experimental data without overfitting. We also used denoising AE in the WoDE case and show that we can significantly reduce the number of training *I*–*V* data to 500 by just varying the parameters of interest (WF and T) despite the existence of other known variations and unknown noise. Our results show that the ML model WoDE can achieve an acceptable performance as the one WDE by predicting the experimental WF and T well.

II. Experimental Data

Fig. 1(a) shows the schematic of the Ga_2O_3 SBDs fabricated on 2-in Ga_2O_3 wafers. These devices are promising candidates for power electronics, RF, and high-temperature sensing applications. The wafer epitaxial structure consists of an n-Ga_2O_3 layer grown on a commercial 2-in n^+-Ga_2O_3 substrate. The substrate has good uniformity in thickness and resistivity. The thickness (t_D) and net donor concentration (N_D) of the n-Ga_2O_3 drift layer were measured at five spots across the wafer using the secondary ion mass spectrometry and electrochemical capacitance–voltage, respectively. A relatively large variation in t_D ($\sim\pm25\%$) and N_D ($\sim\pm35\%$) was found for the epitaxial Ga_2O_3 drift layer.

Field-plated power Ga_2O_3 SBDs were fabricated on the wafer. The anode has diameters ranging from 110 to 260 μm. The details of device fabrication are described in [1], [9], and [10]. Before the Schottky metal deposition, the wafer was cut into small pieces for different surface chemical treatments to intentionally introduce the variations in the Schottky barrier height [11]. Specifically, three chemical treatments were applied, including the deionized (DI) water, hydrochloric (HCl) acid, and dilute hydrofluoric (HF) acid. It was reported that the HCl solutions could remove the surface oxidation layer and produce a lower barrier [11]. The dilute HF treatment introduces fluorine ions in the vicinity of the metal–semiconductor interface, and the negative charge of the fluorine ions [12] raises the barrier. As a result, the Schottky metal has varying effective WF which captures the change of barrier height. The *I*–*V* characteristics of about 20 devices from different pieces were collected, where each device was measured at three chuck temperatures (T), namely Temp1, Temp2, and Temp3. As a result, there are 66 experimental *I*–*V* curves. It is found that the device junction temperature, calibrated by using a thermal couple and a thermal camera when there is no self-heating, depends on the location of the device on a specific sample, the sample size, and the sample shape. Therefore, the measured junction temperature is used in the ML as T with Temp1, Temp2, and Temp3 in the range of 298–308, 348–368, and 403–423 K, respectively.

Fig. 2 shows the measured *I*–*V* of the experimental devices which will be used to deduce T and WF. It can be seen that some device *I*–*V*'s are noisy in the exponential and/or linear regions, which are expected to post challenges to ML if algorithms are not selected carefully because the TCAD trained machine has not encountered the irrelevant variations. Based on the experience in the previous works [1]–[8], full domain expertise is used to extract the most relevant features in the WDE case and the denoising AE is used in the WoDE case.

III. TCAD Data Generation

TCAD Senaturus is used for structure creation [13] and device simulation [14]. Careful model selection and calibration were performed to ensure that the simulation setups capture the correct physics. Philips Unified Mobility Model (PhuMob) [15] is calibrated for Si-doped Ga_2O_3 experimental data [16] to capture the temperature dependence. The incomplete ionization model is turned on due to the ultrawide bandgap in Ga_2O_3. To account for self-heating and model the low thermal resistivity in Ga_2O_3, a half 2-D cross section of the experimental structure is simulated using a cylindrical coordinate, which performs essentially a 3-D simulation of

the SBD in a 0.5 mm^2 sample. Excellent matching between the simulation and experimental data was achieved for one selected device. The calibration details can be found in [1]. A similar setup has also shown to be able to accurately capture the I–V and C–V of an experimental Ga_2O_3 vertical FinFET by another research group [17]. Similar TCAD simulations have been validated as an effective tool to study the electrothermal ruggedness of Ga_2O_3 SBDs in power circuits [18], [19]. Therefore, the model and parameter selections are believed to be accurate enough.

Different from [1], where ~1 50 000 TCAD I–V's are generated, only 500 forward I–V's are used in this study. With <500 curves, the prediction accuracy is found to reduce. Moreover, in [1], four parameters, T, WF, t_D, and N_D, are varied because t_D and N_D are known to have considerable variations in the experimental devices. Therefore, t_D, and N_D, are varied to "teach" the machine the effects of t_D and N_D on I–V's. In this study, only T and WF are varied with fixed t_D and N_D values. T is varied from 300 to 500 K. WF is varied from 4.99 to 5.19 eV. This training framework is to study if the machine would be "smart" enough not to be confused by the relatively large variation in the t_D and N_D values, or in other words, if it recognizes the features in the I–V curves that are only affected by T and WF.

IV. ML WDE

WDE, we can use the knowledge in device physics to find the features in the I–V curves that are only affected by T and WF. Based on device physics, we expect that the SS is related to T. Note that SS is expected to be proportional to kT_{jun}/nq, where k is the Boltzmann's constant, T_{jun} is the junction temperature of the diode, q is the basic electronic charge, and n is the nonideality factor. In the subthreshold region, self-heating is minimal (this is confirmed in the TCAD simulation), so T_{jun} is mainly determined by the chuck temperatures with a minimal impact by device current. Therefore, $T = T_{jun}$ and thus SS is one of the most suitable features to be extracted for the prediction of T. Subsequently, SS between the anode currents of 10^{-5} and 10^{-6} A is extracted as one of the input features.

From device physics, the SBD anode voltage for a given current in the subthreshold region depends on the effective WF of the Schottky metal (or barrier height). Therefore, the anode voltage at which the current is 10^{-5} A (dubbed V_o) is used as the input feature. This captures the most domain expertise with the least effort. V_o and SS form the input features into the machine WDE.

To extract the WF from experimental data, three clean I–V's, measured at the three chuck temperatures (Temp1, Temp2, Temp3), are further calibrated to TCAD simulations (Fig. 3). The calibrated WF at each temperature group is then used as the reference to extract the WF of other I–V's in the same temperature group by finding their relative V_o. For example, if the V_o of a curve in Temp1 is 0.05 V larger than the calibrated Temp1 curve in Fig. 3, its WF is extracted as 0.05 eV + 5.14 eV = 5.19 eV. Here the minor temperature variation, which affects the SS slope, within each temperature group is neglected.

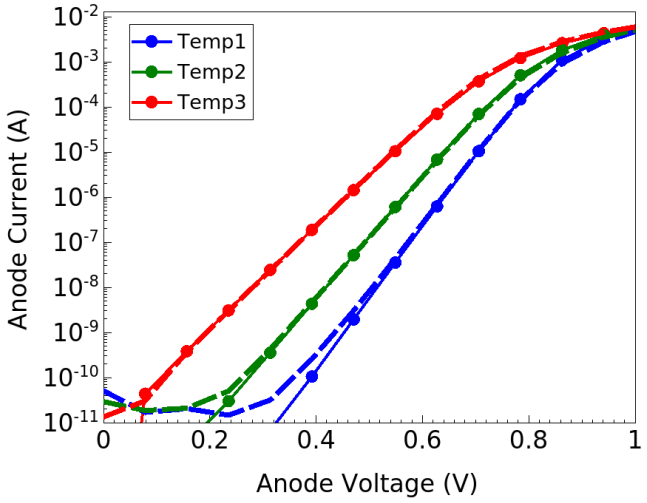

Fig. 3. Calibration of TCAD simulation (lines) to experimental I–V's (dash) from the three temperature groups for WF extraction. To match the experimental curves here (SS), Temp1, Temp2, and Temp3 are found to be 303, 353, and 425 K, respectively. The extracted WFs are 5.14, 5.152, and 5.17 eV, respectively.

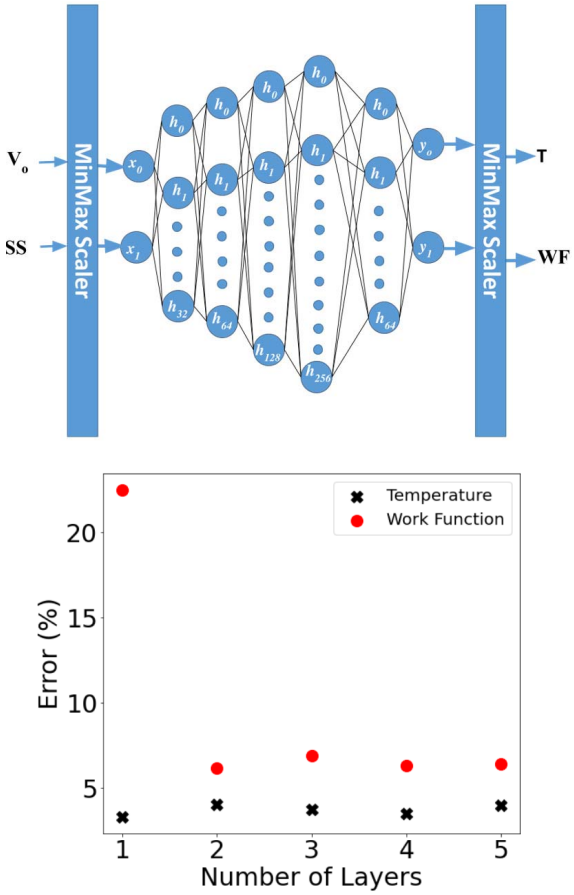

Fig. 4. ML WDE. Top: NN used (five hidden layers showed). Bottom: errors of prediction for different numbers of layers for WDE.

ML WDE algorithm is showed in Fig. 4. V_o and SS are extracted from the TCAD I–V's as the input features. They are scaled using the min-max scaler in scikit-learn [20]. Scaling is necessary because V_o and SS have different orders of

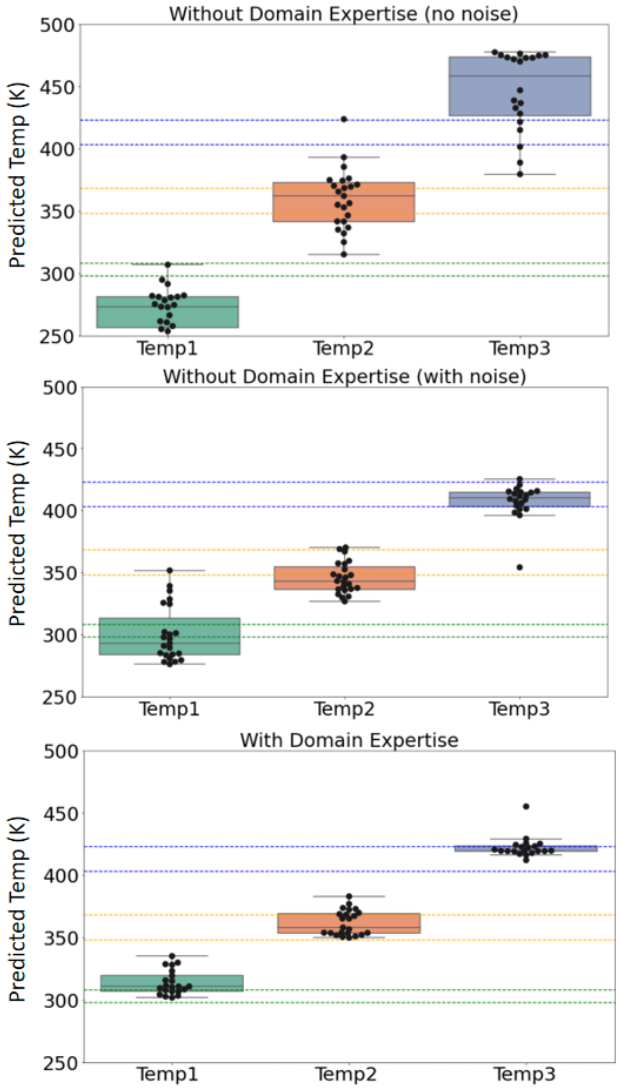

Fig. 5. Seaborn box plot of T prediction by the machine without and with domain expertise. Dotted lines are the expected temperature range of each temperature group.

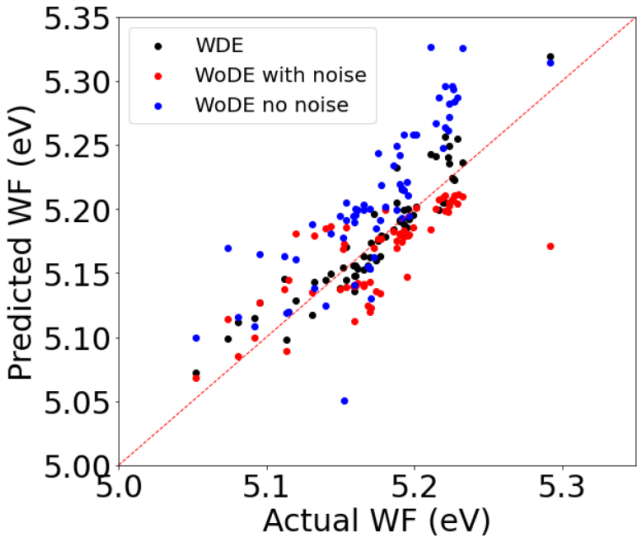

Fig. 6. Scatter plot of WF prediction by the machine with (black) and without domain expertise (red and blue).

magnitude in value. This is also true for T and WF, which have about 100 times difference in magnitude. The neural network (NN) is used to train the machine with T and WF as the outputs through supervised learning. The NN has one input, one output, and five hidden layers. The input and output layers each has two nodes. The activation on the output layer is linear. NN with 1, 2, 3, 4, and 5 hidden layers were tested. For the five-hidden-layer case, the first dense layer has 32 nodes, the second has 64 nodes, the third has 128 nodes, the fourth has 256 nodes, and the fifth has 64 nodes. The hidden layers all use rectified linear unit (ReLU) activation. The hidden layers are removed one by one from the right for 4, 3, 2, and 1-hidden layer cases. Prediction error in percentage is defined as the root-mean-square of the difference between the predicted and expected values, normalized by the experimental variation range. For T, the variation range is 110 K (Fig. 5) and for WF, the range is 110 K (Fig. 6), which is the difference between the means of the expected experimental values in Temp1 and Temp3. It is found that four-hidden-layer NN gives the best result and is used to compare to the WoDE cases. Note that since V_0 depends on WF, T, and also N_D, NN with nonlinear rectification is required.

The trained machine is then used to predict the T's and WF's of the devices based on the V_o and SS extracted from the measured I–V's. Figs. 5 and 6 show the results of T and WF predictions, respectively. As expected, the prediction is very good because the best domain expertise is used. For example, V_0 is just a constant offset from the actual WF. Naturally, the machine can be trained well.

Fig. 4 shows the prediction accuracy of NN with a various number of hidden-layer. It can be seen that two-hidden-layer can achieve good prediction already (while four is the best). Moreover, the accuracy remains low even with five hidden layers. This shows that the extracted features are very relevant that no overfitting occurs even with a large NN.

V. ML WoDE

ML WoDE is then applied to the same data set. The machine is trained by directly using the TCAD I–V curves without any manual feature extraction, and thus no device physics knowledge is required. In [1] and [4], we showed that machines trained using TCAD I–V's WoDE have severe overfitting issues and will predict spurious results on experimental data (e.g., negative temperature and length in kilometer). In [1], PCA is used to alleviate the issue but the temperature predicted is still not good enough (prediction range is 40 K for a given 110 K range data and the values have offset). In [3], noise is added to the data and it gives a range of ~110 K. However, the predicted values are offset and, thus, only the relative values are meaningful.

To combat the overfitting and solve the offset and range issues, here we employed the denoising AE. Fig. 7 shows the algorithm. Each I–V is discretized into 52 points (y_0 to y_{51}). Gaussian white noise with a 7 dB signal to noise ratio (SNR) is added to the TCAD I–V curves (Fig. 8). Their logarithmic values are then taken followed by the min–max scalar.

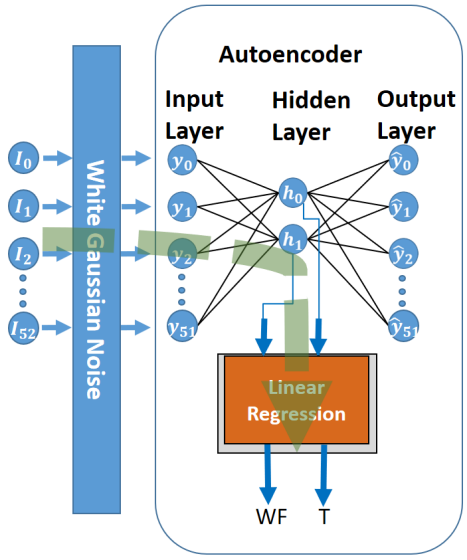

Fig. 7. ML WoDE. The machine is trained such that the output values equals the input values in the AE and h_0 and h_1 are linearly regressed against WF and T.

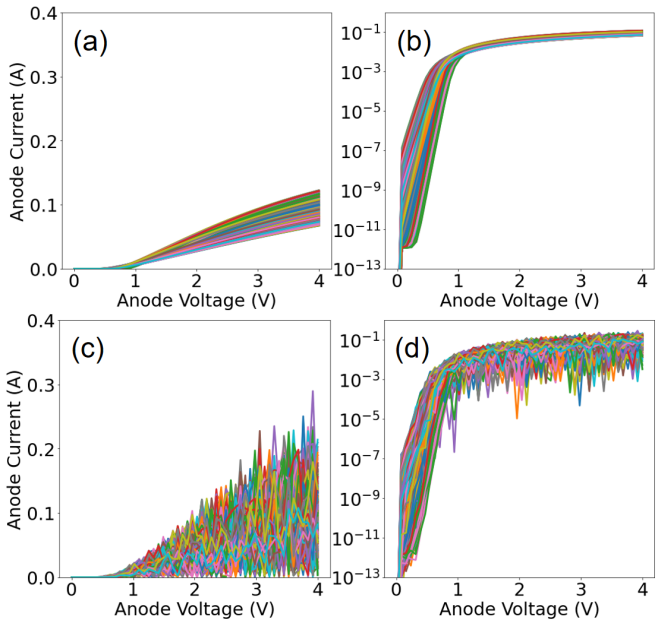

Fig. 8. TCAD I–V's used for training without, (a) and (b) with Gaussian white noise at 7-dB SNR, (c) and (d) in linear, (a) and (c) logarithmic, and (b) and (d) scales.

Note that the AE is trained such that the output I–V equals to the input I–V without noise. The AE has one hidden layer with two nodes. AE can perform manifold learning for dimensionality reduction in hyperspace (e.g., the 52-D space of the I–V curves) to extract the main components (e.g., h_0 and h_1) that are affecting the I–V's [6]–[22]. Since the I–V's are generated by varying the WF and T in TCAD simulations, the hidden nodes (h_0 and h_1) are expected to have correlations with WF and T. Therefore, linear regression is performed to correlate h_0 and h_1 to WF and T.

After the machine is trained, experimental I–V's are then fed into the machine to deduce the WF and T of the device, without any manual feature extraction. Figs. 5 and 6 show the prediction of T and WF by the denoising AE. Its performance is very similar to the one WDE.

TABLE I
PREDICTION ERRORS

	WDE (4 hidden-layer)	WoDE (no noise)	WoDE (with noise)
WF (eV)	6.30%	19.95%	11.88%
T (K)	3.47%	12.36%	5.11%

ML without white noise was also studied and the results are plotted in Figs. 5 and 6 and shown in Table I. The addition of white noise reduces the prediction error of both WF and T by about 50%. Table I summarizes the prediction accuracy.

VI. DISCUSSION

The relatively accurate prediction of experimental data's WF and T by purely TCAD trained machine shows that TCAD trained machine is promising to be used in the semiconductor device and process variation study or reverse engineering. *This is the main breakthrough* in this article because, in most studies, the TCAD trained machines are applied to TCAD data (e.g., [2]–[7], [8]) or did not give satisfactory enough prediction in the experiment data (wrong range or with offset) (e.g., [1]–[3]).

The difficulty of applying TCAD trained machines to experimental data is due to the extra variations in the experimental data. The machines trained by the TCAD data only have "seen" the data due to controlled variations (e.g., WF and T). They are usually overfitted (overconfident) and try to interpret the experimental data variations induced by other sources as the result of the training parameters. This can be naturally solved by using domain expertise to identify the features in the I–V curves. For example, in Fig. 2, since it is identified to extract V_0 and SS in the 10^{-5} to 10^{-6} A region as the features, the variations in the linear region and other SS regions will not affect the performance of the machine. However, this needs too much human intervention.

Therefore, ML WoDE is preferred. This is particularly important if there is no simple relationship between the parameters of interest and the I–V curves. Here we demonstrated denoising AE can achieve this goal by: 1) adding noise to the TCAD data so the machine understands that the I–V can change due to some unknown variations and 2) relying on the AE to perform manifold learning and extract two hidden features (h_0, h_1) from the high dimensional I–V automatically. Usually, we do not understand the meaning of h_0 and h_1. Through linear regression with WF and T, the TCAD trained machine can be applied to the experimental data without being "confused" by the noise and nonideality in the curves.

Adding noise to AE reduces overfitting. However, if the SNR is too small (i.e., too much noise), the signal (i.e., the features of the curves) cannot be learned effectively as the noise is dominating. If the SNR is too large, the noise has less effect. Therefore, it is important to choose an optimal SNR. The SNR chosen here is based on the study in [3]. However, it is not clear how this should be chosen. Fig. 8 shows that with the chosen SNR, the forward current region becomes very noisy. It might thus force the machine to learn from the subthreshold region which has strong features

related to WF and T. Further study is required to understand the mechanism.

VII. Conclusion

In this article, we showed that machines can be generated and trained by TCAD generated data (as few as 500 data) to deduce physical parameters (such as effective WF and ambient temperature) of a given experimental device I–V curve without overfitting. The machine can predict the parameters of interest well despite the existence of unknown noise and known variations of other physical quantities that the machine has not encountered. This can be achieved by two methods. One is using domain expertise to extract the most relevant features from the I–V to train the machine. Another, which is a better and more automatic method, is to use denoising AE to obviate the requirement of domain expertise. Both methods predict the experimental WF and T ranges and trends in an emerging device technology, the ultrawide bandgap Ga_2O_3 SBD, reasonably well. These show the great potential of TCAD-augmented ML in the applications of device and process variation study and reverse engineering.

Acknowledgment

The Virginia Tech authors would like to thank the support from the Jeffress Trust Awards Program in Interdisciplinary Research by Thomas F. and Kate Miller Jeffress Memorial Trust, Bank of America, Trustee.

References

[1] H. Y. Wong *et al.*, "TCAD-machine learning framework for device variation and operating temperature analysis with experimental demonstration," *IEEE J. Electron Devices Soc.*, vol. 8, pp. 992–1000, 2020, doi: 10.1109/JEDS.2020.3024669.

[2] Y. S. Bankapalli and H. Y. Wong, "TCAD augmented machine learning for semiconductor device failure troubleshooting and reverse engineering," in *Proc. Int. Conf. Simul. Semiconductor Processes Devices (SISPAD)*, Udine, Italy, Sep. 2019, pp. 1–4, doi: 10.1109/SISPAD.2019.8870467.

[3] S. S. Raju, B. Wang, K. Mehta, M. Xiao, Y. Zhang, and H.-Y. Wong, "Application of noise to avoid overfitting in TCAD augmented machine learning," in *Proc. Int. Conf. Simul. Semiconductor Processes Devices (SISPAD)*, Kobe, Japan, Sep. 2020, pp. 351–354, doi: 10.23919/SISPAD49475.2020.9241654.

[4] K. Mehta, S. S. Raju, M. Xiao, B. Wang, Y. Zhang, and H. Y. Wong, "Improvement of TCAD augmented machine learning using autoencoder for semiconductor variation identification and inverse design," *IEEE Access*, vol. 8, pp. 143519–143529, 2020, doi: 10.1109/ACCESS.2020.3014470.

[5] C.-W. Teo, K. L. Low, V. Narang, and A. V.-Y. Thean, "TCAD-enabled machine learning defect prediction to accelerate advanced semiconductor device failure analysis," in *Proc. Int. Conf. Simul. Semiconductor Processes Devices (SISPAD)*, Udine, Italy, Sep. 2019, pp. 1–4, doi: 10.1109/SISPAD.2019.8870440.

[6] K. Mehta and H.-Y. Wong, "Prediction of FinFET current-voltage and capacitance-voltage curves using machine learning with autoencoder," *IEEE Electron Device Lett.*, vol. 42, no. 2, pp. 136–139, Feb. 2021, doi: 10.1109/LED.2020.3045064.

[7] J. Chen *et al.*, "Powernet: SOI lateral power device breakdown prediction with deep neural networks," *IEEE Access*, vol. 8, pp. 25372–25382, 2020, doi: 10.1109/ACCESS.2020.2970966.

[8] 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, doi: 10.1109/LED.2019.2931839.

[9] N. Allen *et al.*, "Vertical Ga_2O_3 Schottky Barrier diodes with small-angle beveled field plates: A Baliga's figure-of-merit of 0.6 GW/cm^2," *IEEE Electron Device Lett.*, vol. 40, no. 9, pp. 1399–1402, Sep. 2019, doi: 10.1109/LED.2019.2931697.

[10] B. Wang *et al.*, "High-voltage vertical Ga_2O_3 power rectifiers operational at high temperatures up to 600 k," *Appl. Phys. Lett.*, vol. 115, no. 26, Dec. 2019, Art. no. 263503, doi: 10.1063/1.5132818.

[11] J. Yang, Z. Sparks, F. Ren, S. J. Pearton, and M. Tadjer, "Effect of surface treatments on electrical properties of β-Ga_2O_3," *J. Vac. Sci. Technol. B*, vol. 36, no. 6, Nov. 2018, Art. no. 061201, doi: 10.1116/1.5052229.

[12] Y. Zhang, M. Sun, S. J. Joglekar, T. Fujishima, and T. Palacios, "Threshold voltage control by gate oxide thickness in fluorinated GaN metal-oxide-semiconductor high-electron-mobility transistors," *Appl. Phys. Lett.*, vol. 103, no. 3, Jul. 2013, Art. no. 033524, doi: 10.1063/1.4815923.

[13] *Sentaurus Structure Editor User Guide Version O-2018.06*, Jun. 2018.

[14] *Sentaurus Device User Guide Version O-2018.06*, Jun. 2018.

[15] D. B. M. Klaassen, "A unified mobility model for device simulation-I. Model equations and concentration dependence," *Solid-State Electron.*, vol. 35, no. 7, pp. 953–959, Jul. 1992, doi: 10.1016/0038-1101(92)90325-7.

[16] K. Goto *et al.*, "Halide vapor phase epitaxy of Si doped β-Ga_2O_3 and its electrical properties," *Thin Solid Films*, vol. 666, pp. 182–184, Nov. 2018, doi: 10.1016/j.tsf.2018.09.006.

[17] H. Y. Wong and A. C. F. Tenkeu, "Advanced TCAD simulation and calibration of gallium oxide vertical transistor," *ECS J. Solid State Sci. Technol.*, vol. 9, no. 3, Feb. 2020, Art. no. 035003.

[18] C. Buttay, H.-Y. Wong, B. Wang, M. Xiao, C. Dimarino, and Y. Zhang, "Surge current capability of ultra-wide-bandgap Ga_2O_3 Schottky diodes," *Microelectron. Rel.*, vol. 114, Nov. 2020, Art. no. 113743, doi: 10.1016/j.microrel.2020.113743.

[19] M. Xiao *et al.*, "Packaged Ga_2O_3 Schottky rectifiers with over 60 a surge current capability," *IEEE Trans. Power Electron.*, early access, Jan. 8, 2021, doi: 10.1109/TPEL.2021.3049966.

[20] *Scikit-Learn*. Accessed: Jan. 1, 2021. [Online]. Available: https://scikit-learn.org/stable/

[21] G. E. Hinton, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006, doi: 10.1126/science.1127647.

[22] M. A. Kramer, "Nonlinear principal component analysis using autoassociative neural networks," *AIChE J.*, vol. 37, no. 2, pp. 233–243, Feb. 1991, doi: 10.1002/aic.690370209.