IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 30 January 2023, accepted 12 February 2023, date of publication 14 February 2023, date of current version 21 February 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3245525

APPLIED RESEARCH

Deep Neural Networks for Determining Subgap States of Oxide Thin-Film Transistors

YUNYEONG CHOI[1], (Member, IEEE), WOOKYUNG SUN[2], JISUN PARK[1,3], AND HYUNGSOON SHIN[1,3], (Senior Member, IEEE)

[1]Department of Electronic and Electrical Engineering, Ewha Womans University, Seoul 03760, South Korea
[2]Department of Electrical and Computer Engineering, Seoul National University, Seoul 08826, South Korea
[3]Graduate Program in Smart Factory, Ewha Womans University, Seoul 03760, South Korea

Corresponding authors: Jisun Park (jisunpark@ewha.ac.kr) and Hyungsoon Shin (hsshin@ewha.ac.kr)

This work was supported by the National Research Foundation (NRF), South Korea, under Project BK21 FOUR.

ABSTRACT In this study, we propose a deep neural network (DNN) model that extracts the subgap states in the channel layer of oxide thin-film transistors. We have developed a framework that includes creating a model training set, preprocessing the data, optimizing the model structure, decoding from density-of-state (DOS) parameters to current–voltage (I–V) characteristics, and evaluating the model performance in terms of curve fitting accuracy. We investigate in detail the effect of data preprocessing methods and model structure on the performance of the model. The primary finding is that the input data type and the last hidden layer significantly affect the performance of the regression model. Using double-type input data composed of several voltages and linear current values is more advantageous than using log-scale current. Moreover, the number of nodes in the last hidden layer of a regression model with multiple output nodes should be large enough to avoid interference between the output values. The proposed model outputs five DOS parameters, and the resulting parameters are decoded to an I–V curve through interpolation based on the nearest 32 data from the given dataset. We evaluate the model performance using the threshold voltage and on-current difference between a target curve and the decoded curve. The proposed model calibrates 97.1% of the 14,400 curves within the threshold voltage difference of 0.2V and on-current error of 5%. Hence, the proposed model is verified to effectively extract DOS parameters with high accuracy based on the current characteristics of oxide thin-film transistors. We expect to improve the efficiency of defect analysis by replacing the iterative manual technology computer aided design (TCAD) curve fitting with an automatic DNN model.

INDEX TERMS Deep neural network (DNN), thin-film transistor (TFT), defect, subgap state, density of states (DOS), TCAD, supervised learning, regression model, preprocessing, input data type, hidden layer, model structure.

I. INTRODUCTION

Oxide semiconductor thin-film transistors (hereafter, oxide TFTs) have considerable potential in electronics applications such as displays, sensors, memory, Internet of things (IoT), energy-harvesting, and medical/bio-interface devices [1], [2]. Oxide TFTs have been employed since 2012 specifically for flat-panel displays owing to their excellent electrical properties such as high TFT mobility >10 cm^2/Vs and low leakage current. Moreover, oxide TFTs have certain advantages for flexible and large-scale display applications, including low process temperature which is compatible with plastic substrates, and scalability owing to the high mobility even in an amorphous phase [3], [4], [5], [6]. In amorphous oxide semiconductors (AOS), the metal and oxygen ions create a Madelung potential that stabilizes the ionized states with a bandgap of about 3 eV. A bandgap is the energy difference between the top of the valence band (E_V), which is mostly full of electrons, and the bottom of the conduction band (E_C), which is mostly void of electrons. Conduction band minimums (CBM) and valence band maximums (VBM) are

The associate editor coordinating the review of this manuscript and approving it for publication was Chaitanya U. Kshirsagar.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

mainly made of empty metal cation orbitals and fully occupied O 2p orbitals, respectively [5].

Ideally, no trap states are allowed in the bandgap. However, in reality, there are subgap states that originate from atomic disorders such as oxygen vacancies and interstitials [7]. There are two types of defects depending on the electrical properties—acceptor-like and donor-like states. Acceptor-like states below the Fermi level (E_f) capture electrons, reducing the current, while donor-like states above the E_f donate electrons, increasing the current. Equations 1–5 define the distribution of subgap states. The total of the subgap states, $g(E)$, is composed of four bands: two tail bands, acceptor-like tail states near the E_C and donor-like tail states near the valence band), and two deep-level bands (one acceptor-like and the other donor-like), which are modeled using a Gaussian distribution [8]. Ten parameters determine the DOS distribution: N_{TA}, W_{TA}, N_{GA}, E_{GA}, W_{GA}, N_{TD}, W_{TD}, N_{GD}, E_{GD}, and W_{GD}. The subscripts T, G, A, and D represent the tail, Gaussian (deep level), acceptor, and donor states, respectively.

$$\begin{aligned} g(E) &= g_{TA}(E) + g_{TD}(E) + g_{GA}(E) + g_{GD}(E) \\ &g(E) : total\ density-of-state\ (DOS) \\ &g_{TA}(E) : acceptor-like\ tail\ states \\ &g_{TD}(E) : donor-like\ tail\ states \\ &g_{GA}(E) : acceptor-like\ Gaussian\ state \\ &g_{GD}(E) : donor-like\ Gaussian\ state \end{aligned} \tag{1}$$

$$g_{TA}(E) = \mathrm{N_{TA}} exp\left[\frac{E - E_c}{\mathrm{W_{TA}}}\right] \tag{2}$$

$$g_{TD}(E) = \mathrm{N_{TD}} exp\left[\frac{E_v - E}{\mathrm{W_{TD}}}\right] \tag{3}$$

$$g_{GA}(E) = \mathrm{N_{GA}} exp\left[-\left[\frac{\mathrm{E_{GA}} - E}{\mathrm{W_{GA}}}\right]^2\right] \tag{4}$$

$$g_{GD}(E) = \mathrm{N_{GD}} exp\left[-\left[\frac{E - \mathrm{E_{GD}}}{\mathrm{W_{GD}}}\right]^2\right] \tag{5}$$

The equations and DOS parameters only determine the DOS distribution of subgap states inside the oxide semiconductor and are not directly related to the internal function of the deep learning model. The proposed model learns the correlation between the DOS parameters and the I-V information without considering device physics. Current-voltage characteristics dependent on DOS parameters are obtained through technology computer-aided design (TCAD) simulation. As the E_f of an AOS TFT in an operating range is near the E_C, acceptor-like tail states and donor-like Gaussian states have dominant effects on current-voltage characteristics. Therefore, this paper focuses on $g_{TA}(E)$ and $g_{GD}(E)$ among the four bands. The density of acceptor-like tail states defined by N_{TA} is highest at E_C and decreases with a slope of $1/W_{TA}$ as the energy level decreases. The peak density of donor-like Gaussian states is defined by N_{GD}, and the mean energy level is 0.1–0.5 eV below the E_C with a characteristic decaying energy of W_{GD} (Fig. 1).

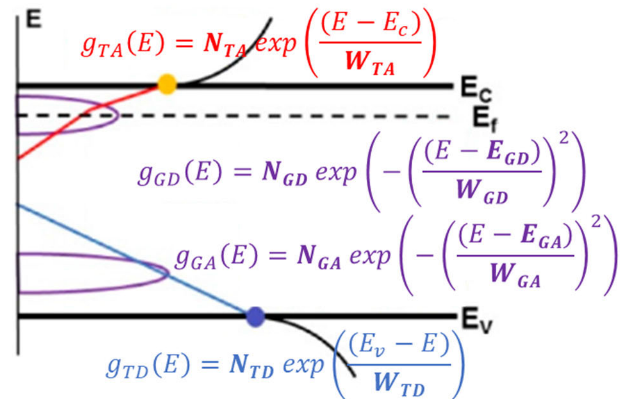


FIGURE 1. Graph of density-of states as a function of energy and distribution modeling equations with DOS parameters.

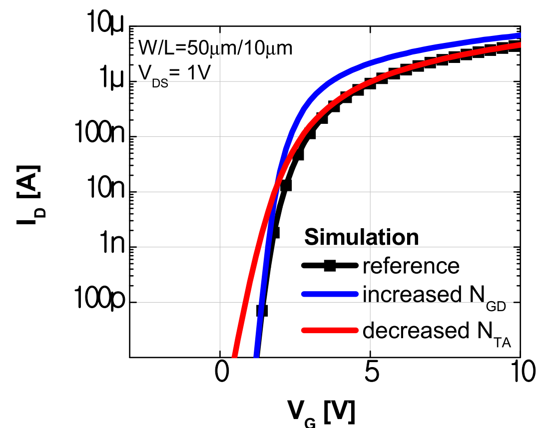


FIGURE 2. Comparison of I-V characteristics with different DOS parameters of increased NGD or decreased N_{TA}.

The subgap states affect the electrical characteristics of AOS TFTs [9], [10], [11]. Generally, when donor-like Gaussian states increase, the number of donated electrons from the trap states increases, and the threshold voltage of a device is negatively shifted, which means a higher current flows in the device at the same voltage condition. On the other hand, when acceptor-like tail states decrease, the traps capture the electrons contributing to the device current, resulting in an increase of on-current (Fig. 2). Even devices with the same structure have different threshold voltage (V_{TH}), subthreshold swing (SS), and on-current level depending on the density of states (DOS). Thus, determining the subgap states of a channel layer is essential to analyze the characteristics of oxide TFTs.

Several researchers have studied the subgap states of oxide TFTs for various processes and stress conditions and have characterized the distribution of donor-like state changes. [12], [13], [14]. In the case of illumination bias instability, it has been surmised that electrons exit from a deep-level donor state when exposed to light and cause a negative V_{TH} shift [7], [15], [16]. The DOS change due to the re-arrangement of atoms in oxide semiconductors accounts for device degradation under bending stress [17], [18], [19], [20]. There are several methods for the experimental

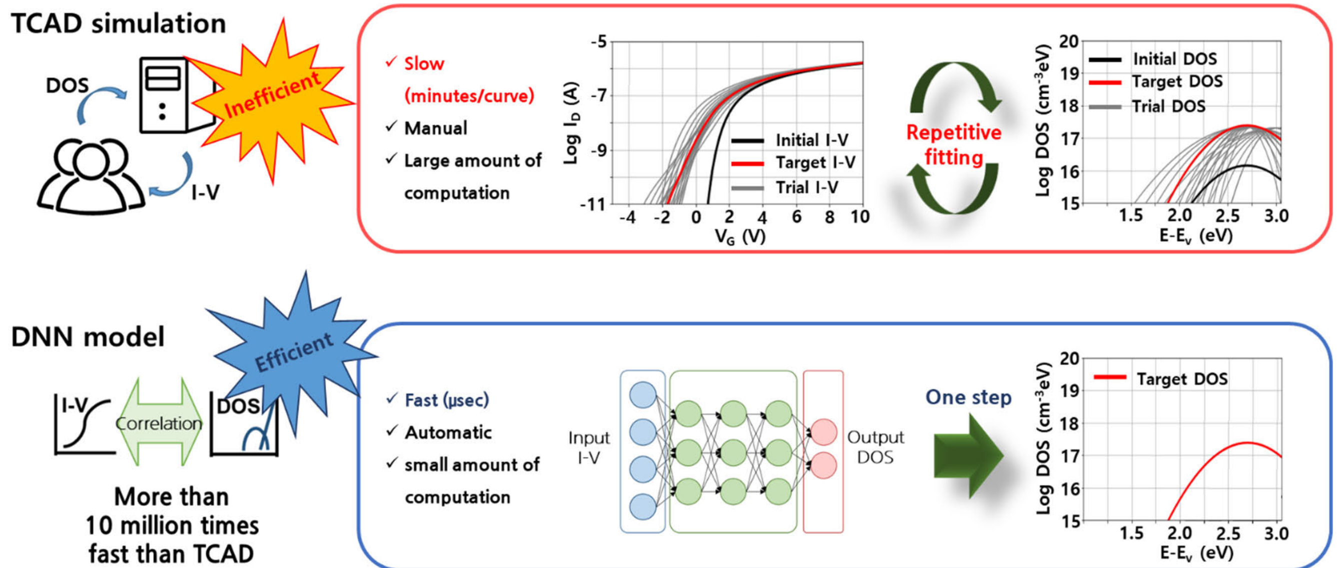


FIGURE 3. Comparison of curve fitting for determining subgap state of a TFT using (a) TCAD simulation and (b) DNN model. The total time consumption of TCAD simulation is multiplied by the number of iterations.

detection of trap densities, such as electron paramagnetic resonance, deep-level transient spectroscopy, isothermal capacitance transient spectroscopy, constant photocurrent methods, device simulation fitting, field-effect (FE) methods, and capacitance–voltage (C–V) methods [21]. Although experimental measurement methods give detailed information on trap distribution, separate equipment is required to detect the traps in different energy ranges. The curve fitting method is widely used because it does not require additional measurement and enables various simulations using the extracted DOS by fitting [22], [23], [24].

However, the curve fitting method involves an iterative process of optimizing DOS parameters until the simulation curve fits the measurement curve well (Fig. 3(a)). The parameters are decided manually in this process, and substantial computational power is required for the TCAD simulation. Determining an appropriate estimate of the DOS value requires considerable experience and repetitive simulations. We propose a deep learning model for automatic extraction that can replace the time-consuming manual process (Fig. 3(b)).

A classification model targets qualitative data to predict the class to which a sample belongs, and a regression model outputs quantitative data to fit a specific target value. Attempts have been made to analyze transistor characteristics using ML. In 1996, an early ML model by Meijer was proposed for predicting an I–V curve using an artificial neural network composed of 12 neurons [25]. The model structure was simple due to the limitations of computational memory, and the model could not sufficiently learn the complex information. Recent models based on the gate and drain voltages have been presented for predicting the current value [26], [27], [28], [29]. These models are proposed to replace compact models in circuit simulation. Since these models aimed to calibrate a particular I–V characteristic, they could not be applied to analyze characteristics with a wide range of deviations.

Studies on transistor defects have been also presented. Teo et al. suggested a model for predicting the defect location of single-fin FinFETs based on I–V information using a random forest algorithm [30]. Kocak et al. presented a model for determining the presence of defects in devices based on gate, source, drain, and substrate current graph images using the convolutional neural network (CNN) [31]. However, because this classification model predicts one class out of 2 or 10 labels, it is different from a regression model that needs to calibrate accurate DOS values. A model that predicts the threshold voltage, on-current, off-current, SS, and drain-induced barrier lowering (DIBL) based on the coordinates, drain voltage, and trap energy level of a single trap of a bulk FinFET using a gradient boosting decision tree was also suggested [32]. However, it could not be used for our purpose as the model and data type were different.

There are regression models that predict the threshold voltage or SS based on the structural information regarding the device [33], [34], [35]. However, they could not be applied to our model for three reasons. First, the data trends were significantly different from our model, which predicts DOS based on I–V characteristics. Second, our model has multiple output nodes, and DOS parameters cannot be extracted individually because of the correlation between parameters. Third, the accuracy of the model varies depending on the number of nodes in the output stage. Therefore, it is necessary to develop a data preprocessing technique and a neural network model that considers data characteristics for the DOS analysis.

In this paper, we propose a model that can quickly and automatically extract the subgap state of a TFT based on I–V characteristics. We also present the guidelines for data preprocessing techniques and model structures. Most studies

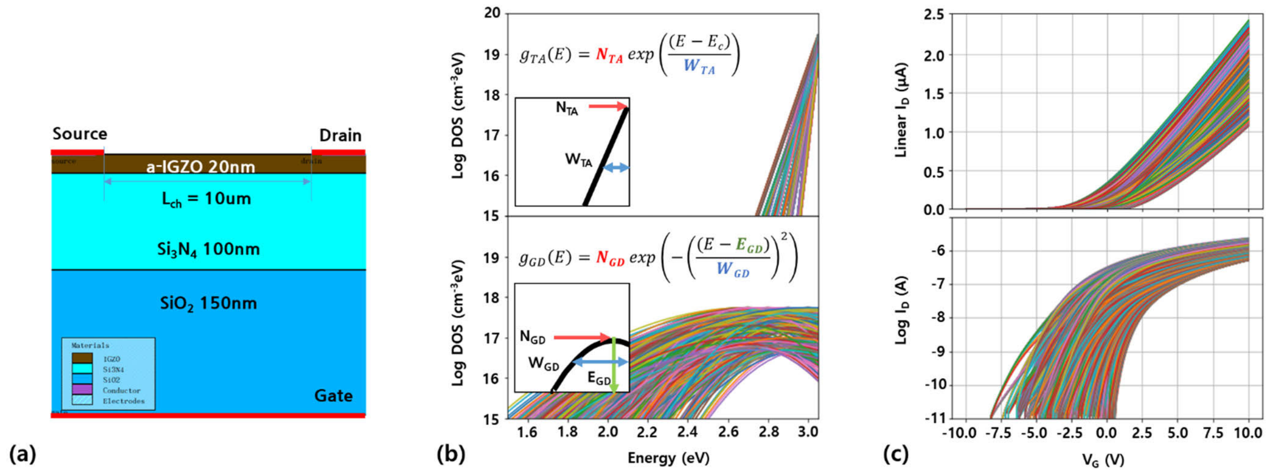


FIGURE 4. TCAD simulation condition and result. (a) Simulation structure, (b) subgap state distribution applied to the channel layer with 14,400 combinations of five DOS parameters resulting in (c) 14,400 I–V characteristics with a large variation of V_{TH}, SS, on-current.

about deep learning for device characterization use I–V characteristics without delicate preprocessing. CNN models utilize I–V graph images to provide current and voltage information simultaneously, and deep neural network (DNN) models assign all current measurements to the input nodes in order after a logarithmic operation. However, this method is inefficient because of its low performance and complexity. Therefore, a model for DOS extraction is required. In this study, we tested four data types for input data, namely linear, logarithmic currents, voltages, and both current and voltage, to suggest the most effective data preprocessing method that covers an entire operating region. In addition, model performances are compared for a different number of hidden layers and nodes.

II. METHOD

The development process of a model comprises three steps: creating a dataset, training the model, and evaluating curve fitting accuracy with a test set. First, the dataset for ML is obtained by TCAD simulation. The method of preprocessing is focused to improve the model performance. After preprocessing, in the training phase, the model learns to minimize the mean square error (MSE) loss of DOS by an iterative update of the weights. We present two models according to input type and experiment with various hidden layer compositions to find the dominant elements of the regression model. Lastly, in the testing phase, the model performance is evaluated by the curve fitting accuracy. After the model outputs DOS parameters using test input data, the parameters are converted back to an I–V curve by the decoding method using interpolation. The accuracy of the model is calculated based on whether the decoded curve fits well with the test input I–V curve. The detailed steps are described as follows.

A. MACHINE LEARNING DATASET

Figure 4(a) shows a device structure used in the simulation with a channel width and length of 10 μm. Various I–V curves are obtained by applying 14,400 combinations of subgap state parameters to the channel layer having the same structure (Figure 4(b)). Because the fermi level of the oxide TFT is near the conduction band in the operating range, shallow donor-like states and acceptor-like states near E_c have a dominant effect on the device characteristics. Thus, we split the three parameters of donor-like Gaussian states $g_{GD}(E)$, N_{GD}, E_{GD}, and W_{GD}, and two parameters of acceptor-like tail states $g_{TA}(E)$, N_{TA} and W_{TA}. N_{GA}, E_{GA}, W_{GA}, N_{TD}, and W_{TD} are constant and are 1.0 x 10^{17} (cm^{-3}), 1.0 (eV), 0.50, 1.0 x 10^{19} (cm^{-3}), and 0.05, respectively. The rest of the DOS parameters, N_{GA}, E_{GA}, W_{GA}, N_{TD}, and W_{TD}, are constant and are 1.0 x 10^{17} (cm^{-3}), 1.0 (eV), 0.50, 1.0 x 10^{19} (cm^{-3}), and 0.05, respectively. The N_{GD}, E_{GD}, W_{GD}, N_{TA}, and W_{TA} ranges are 5.0 $\times$ 10^{16}–5.5 x 10^{17} (cm^{-3}), 2.7–3.0 (eV), 0.15–0.50, 5.0 $\times$ 10^{18}–2.0 x 10^{19} (cm^{-3}), and 0.015–0.030, respectively. The total number of I–V curves is 14,400 with 15 N_{GD}s (0.05–0.1 order intervals), 6 E_{GD}s (0.06 intervals), 8 W_{GD}s (0.05 intervals), 8 N_{TA}s (0.1 order intervals), 5 W_{TA}s (0.005 intervals).

Figure 4(c) shows the I–V characteristics. Considering the gate voltage resulting in a current of 0.1 nA as a V_{TH}, the variation of V_{TH} ranges from −7 V to +1 V. Because the range of V_{TH} is wide, it is advantageous to group together samples with similar characteristics and train them separately rather than training the entire dataset with one artificial neural network at a time. The samples were categorized into nine groups based on SS. In the subthreshold region, the current in log scale and gate voltage have a linear relationship, and SS is calculated as a reciprocal of the slope, which is the voltage difference required to increase one order of current. In our paper, we calculate SS as the difference in the gate voltage at I_D = 1 nA and 10 nA. Grouping a dataset with similar samples improves model performance. The dataset is divided by SS, as it enables samples with different trends to be easily separated. The impact of E_{GD} and W_{GD} on the I-V curves depends on N_{GD}. When N_{GD} has a low value, the curves have

steep slopes, and variations of E_{GD} and W_{GD} barely change the slopes. However, when N_{GD} has a high value, E_{GD} and W_{GD} have significant effects on the curves, especially SS. Increasing N_{GD} has a consistent tendency to decrease V_{TH}, regardless of the rest of the DOS parameters. Meanwhile, E_{GD} and W_{GD}, which determine the energy level and shape of donor-like states, have different effects on the curves depending on the level of N_{GD}. Since the effects of E_{GD} and W_{GD} do not have consistent trends, they become factors that degrade model performance. By classifying the dataset based on SS, it is possible to group samples according to the impact of E_{GD} and W_{GD}, which improves model performance.

Because the dataset is divided into several groups, the number of data in a group is lower than that in the total dataset. The data was augmented using Gaussian noise to prevent overfitting. A TCAD-simulated I–V curve consists of 101 current information under a gate voltage range from −10 V to +10 V (0.2 V step). By adding Gaussian noise with a standard deviation of 0.01 for the current at each point, we have made at least 20,000 curves for each group. The DOS parameters for the I–V curve adding the current noise are the same as before data augmentation. The training and test sets are separated using the K-fold method, and the number of folds is ten.

B. DATA PREPROCESSING

In this study, four input data types were considered: linear current type, log current type, voltage type, and current–voltage double-type. Input nodes were sequentially assigned a current value when current-type input data was used. For example, the first, second, and third input nodes had I_D values at $V_G = -10, -9, -8$ V in order, and the last node had an I_D value at $V_G = 10$ V. The gate voltage was swept from −10 V to 10 V in steps of 0.2 V and the total current points were 101.

When the log-current-type input data was used, all 101 current points were used. Since the current level in the sub-threshold region was several orders of magnitude smaller than the one in the on-region, $log_{10}(I_D)$ operation was preceded to utilize the data for the entire operation region. The noise current values under 0.1 pA in the turn-off region were converted to zero. The linear-current-type input data was similar to the log-current-type input data; however, only five points in the on-region were used in the linear scale.

In the voltage input type, the input node received a voltage value required to achieve a specific flow of current. For example, the first, second, and third nodes had the V_G values flowing 0.01 nA, 0.1 nA, and 1 nA, and the last node had V_G at $I_D = 1\mu$A. The accurate voltage information for the specific current level cannot be found directly from the measurement because I–V curves are measured by sweeping V_G in general. Therefore, an additional calculation is required to determine the voltage at a specific current level based on the measurement near the sampled current.

Lastly, the current–voltage double-type input data, which used both linear currents and voltages, covered the entire operating range using only 10 to 12 points. Current and voltage type data reflect the linear and sub-threshold region, respectively. As will be mentioned in the result section, utilizing double-type data is the most effective among the four types.

To improve the performance of the model, preprocessing of output data is also required because N_{GD}, E_{GD}, and W_{GD} have different ranges of values. For N_{GD} and N_{TA}, the density of the defect states was over 1×10^{16} and had a value between 16 and 19 after log operation. E_{GD} and W_{GD} ranged 2.7–3.0 and 0.15–0.50, respectively. There were deviations of one to two orders of magnitude in the parameters. When the loss was calculated by the joint MSE of five parameters, the model was trained to lower the N_{GD} or N_{TA} error in the iterative learning process. Therefore, the target data were normalized to 0–1, considering the minimum and maximum of each parameter.

Moreover, input data validation was conducted by checking abnormal data in the input data. First, we confirmed whether all voltage-type input data were valid and error free due to the conversion process from current to voltage value required to flow a specific reference current. When the reference current is too high, some samples include abnormal data because their source data does not contain information near the reference current level. To prevent erroneous cases, we limit the maximum reference current to the current value of a device with the lowest on-current. Second, we confirmed performance degradation due to Gaussian noise. In our paper, data augmentation using Gaussian noise is conducted to increase the size of the training set. When the standard deviation of the Gaussian noise is too large, the data accuracy decreases, which degrades the model's performance.

C. MODEL STRUCTURE AND FUNCTIONS

Figure 5 shows the proposed DNN model. The input layer consists of 5–101 nodes depending on the input data type. The hidden layer consists of four layers composed of 32, 32, 32, and 64 nodes. A model structure using one input data type among current and voltage is fully connected (Figure 5(a)–(c)). The only difference is the number of input layers and input data types. Meanwhile, a model for the double-type data (current–voltage) is connected only partially (Figure 5(d)). The hidden layers in the front are separated for different types of input data and merged near the last hidden layer. The total number of nodes is the same for all the cases. Table 1 summarizes the model structure.

The output layer consists of five nodes corresponding to N_{GD}, E_{GD}, W_{GD}, N_{TA}, and W_{TA}. These DOS parameters cannot be extracted because they are correlated with each other. For example, when N_{GD} is small, the curve is determined by N_{GD}, N_{TA}, and W_{TA} regardless of E_{GD} and W_{GD}; however, when N_{GD} is large, the curve characteristics change depending on E_{GD} and W_{GD}. In addition, the donor- and acceptor-like states have a considerable effect on the electrical characteristics. When the density of the donor-like state exceeds 10^{17} cm^{-3}, the on-current also increases. When the density of the acceptor-like states is high or the distribution

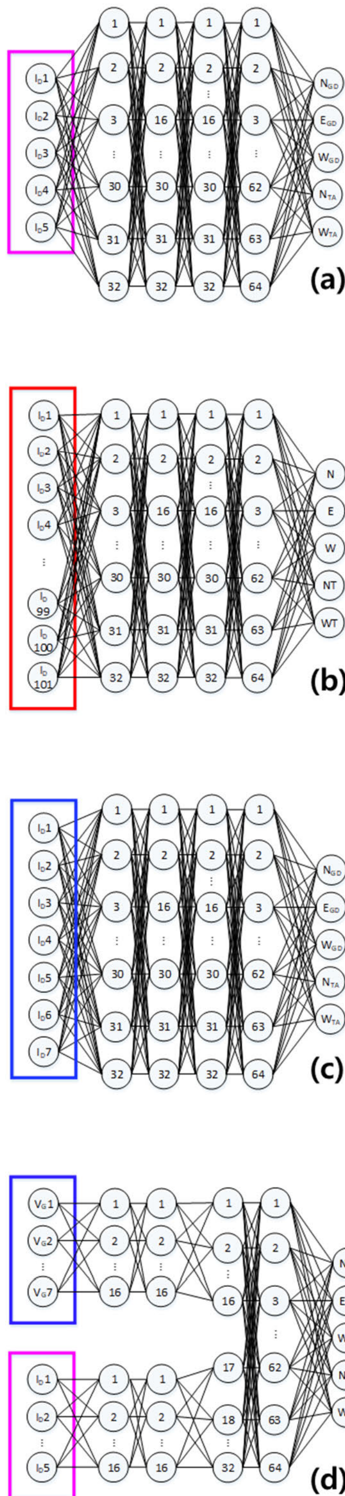


FIGURE 5. Model structure with four types of input data. Models for single-type of input data consists fully connected layers, (a) drain current in a linear scale (b) log scale (c) gate voltage. A model structure for (d) the double-type data consists of a separated and merged parts.

decays slowly, V_{TH} and SS are also affected. The regions affected by each type of DOS are not clearly divided and these DOS parameters need to be considered together. Therefore, five parameters having five nodes in the output layer are estimated simultaneously.

A value of the previous stage is only forwarded without recursive operation, and the activation function is the rectified linear unit (ReLU). Although other activation functions have also been checked, the ones that can have a linear value show better performance. Loss is calculated by the average joint MSE of parameters between the targets and results in a mini-batch. An Adam optimizer updates the model with a learning rate of 0.001. The mini-batch size is 200, and there are 30,000 training iterations.

D. EVALUATION

The goal of this model is to replace the curve fitting process using TCAD, which is usually performed manually and is time-consuming. It is also important to fit curves well to reduce DOS errors. After the deep learning model outputs the DOS parameters, it is necessary to convert the parameters back to I-V curves and compare them with input curves for model validation. However, if the TCAD simulation must be used in the validation process, the advantage of using artificial neural network models is limited. Therefore, we suggested the curve decoding method based on the given dataset without using TCAD simulation.

If the two known points are given by the coordinates (x_0,y_0) and (x_1,y_1), the linear interpolant is the straight line between these points (Figure 6). For a value x in the interval (x_0, x_1), the value y along the straight line is given from the equation of the slopes. In this case, we simply set E_{GD} obtained by the DNN model as x assuming that x_0 and x_1 are the *n*th and *(n+1)*th E_{GD} ($E_{GD}[n]$, $E_{GD}[n+1]$), and y_0 and y_1 are a current value at a specific V_G corresponding to $E_{GD}[n]$, $E_{GD}[n+1]$. For example, when I-V curves of $E_{GD} = 2.7$ (Curve A) and 2.8 (Curve B) are known and the curve of $E_{GD} = 2.74$ (Curve C) is needed for evaluation, the unknown current values of Curve C can be calculated by Equation 6. The distance between E_{GD} values determines the weight of the known curves. In Equation 6, x and y denote E_{GD} and the current value with subscripts 0 and 1 representing Curves A and B, respectively. The maximum distance of the nearest DOS parameter, x1-x0, is 1 in this example. The weight of the current value of Curve A is the difference between x0 and x over the maximum distance, which is 0.6, while that of Curve B is 0.4. Since there are 101 points of current values in the gate voltage range from −10 V to 10 V with an interval of 0.2 V, the interpolation process should be repeated 101 times to decode Curve C.

Assuming all other DOS parameters of curve A and curve B are the same, the current value at V_G can be calculated from the other two known points by Equation (6). As our model has five DOS parameters, there are $32(= 2^5)$ nearest DOS combinations, and a decoded current value can be calculated through 31 interpolations.

$$y = \frac{y_0\,(x_1 - x) + y_1\,(x - x_0)}{x_1 - x_0} \tag{6}$$

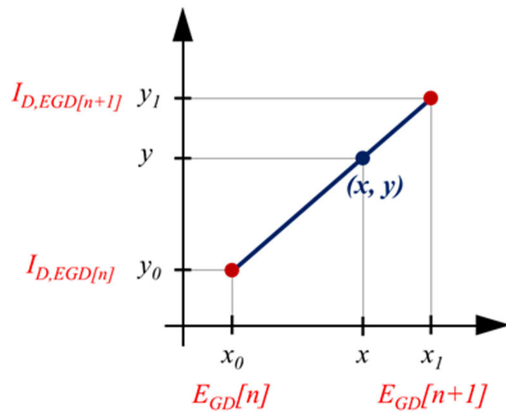


FIGURE 6. Current decoding method using Interpolation between two adjacent information.

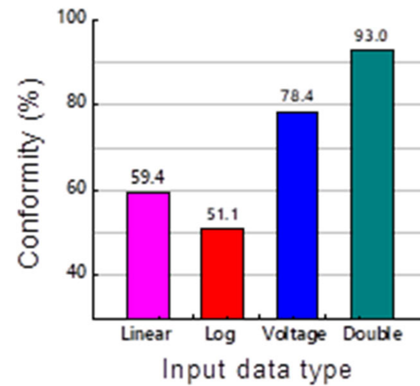


FIGURE 7. Comparison of performance in four cases with different input data types.

This decoding process is 1000 times faster than TCAD simulation, and the accuracy is also adaptable. In the evaluation step, we check the threshold voltage differences (V) and on-current differences (%) between the target curves and result curves corresponding to a DOS extracted by the DNN model.

III. RESULT

We compared the performance depending on the input data type and model structure. The used model structure is summarized in Table 1. The ratio of the curve among 14400 test curves meeting a criterion was expressed as conformity (%). The criteria allow a threshold voltage difference of under 0.2 V and on-current errors under 5%. The time taken to analyze all 14400 test samples using the DNN model is approximately 1.15 s, which is approximately 40 μs per curve.

A. INPUT DATA TYPE

Figure 7 compares the model performance based on various input data types, linear current type, log current type, voltage type, and double type. The log current type utilized current values of 101 points and other types used only 5 to 12 points of input information. Counterintuitively, the log current data type that provides the largest quantity of information is associated with the lowest performance. This implies that providing a small quantity of refined information is more advantageous than simply providing a large amount of data.

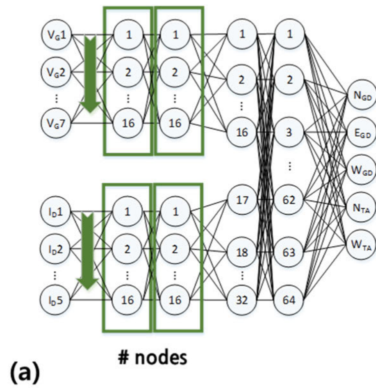
Performance depends on the input data type because each data type covers a different operating range. First, in the case of the linear current type, only the value in the on-region was considered because the current level of the subthreshold region was five to six orders of magnitude lower than the on-current. There was little difference in the input data between samples having similar on-current but different V_{TH} and SS. When the target data was different, but the input data was the same, an ML model tended to output the average of the target by trying to minimize loss regardless of the input data. Thus, the limited availability of linear scale information degraded the performance of the model.

Second, the log current type covers the entire operating range, but it has an additional error factor due to noise filtering. In this type, all current values less than 10^{-12}A were converted to zero and the voltage at which conversion starts depends on the V_{TH} of a device. For example, in a device with a threshold voltage of 0V, half of the information of 101 points was filtered out. The problem was that step patterns were undesirably included in the input data as the number of filtered points varied discretely. Performance is better in the case of noise filtering than without noise filtering, but log current type data also adds error elements during the preprocessing.

The third input data type was voltage. This type has the advantage of being able to reflect subthreshold characteristics regardless of the threshold voltage of the device. However, the information area was limited below the lowest on-current among all samples, because conversion from a current to a voltage by interpolation was possible when the corresponding current flowed through all the devices. The model performance deteriorated, as on-region information was lost.

Finally, the current–voltage double-type covered the entire operating range without noise current using both linear current and voltage. Each data type was assigned to the upper and lower part of the first hidden layer and was forwarded separately. The separated hidden layers were merged in the last hidden layer. The double-type shows the highest performance among the four input types taking advantage of each type.

It is concluded that the pre-processing method is important to provide compact information covering the entire operation range for improving the model accuracy. I-V characteristics can be sufficiently expressed with only about 10 pieces of information. When using a dataset with a wide range of threshold voltages, it is more advantageous to use the data converted to voltage than to use the log current, because voltage-type data can cover the information in the subthreshold region excluding noise without additional error factors. Also, performance can be improved by two types of information covering the entire region.



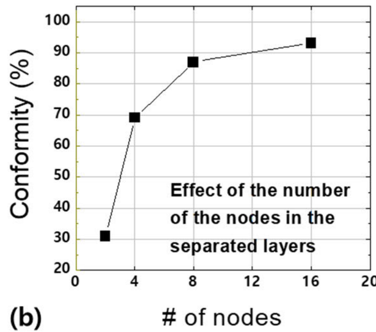
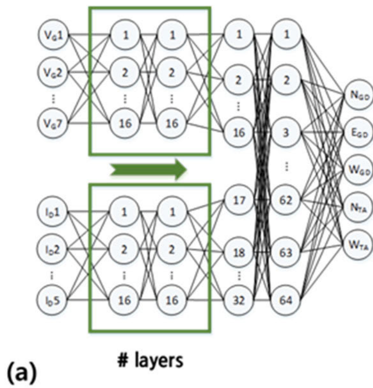


FIGURE 8. Effect of increasing the node in the separated part. (a) Model structure increasing width of the separated part and (b) performance. Each current or voltage type input data is forwarded through 2 layers with 2, 4, 8, and 16 nodes before merging.



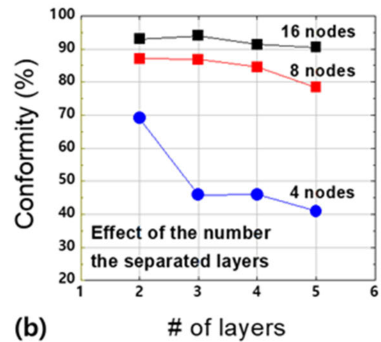


FIGURE 9. Effect of increasing the number of layers in the separated part. (a) Model structure increasing depth of the separated part and (b) performance. The layer has 4 nodes (blue), 8 nodes (red), 16 nodes (black).

TABLE 1. Comparison of model performance according to the number of nodes in the separated part.

Separated part		Merged part		Conformity (%)
# of nodes	# of layers	# of nodes	# of layers	
2	2	32	1	30.9
4	2	32	1	69.1
8	2	32	1	87.0
16	2	32	1	93.0

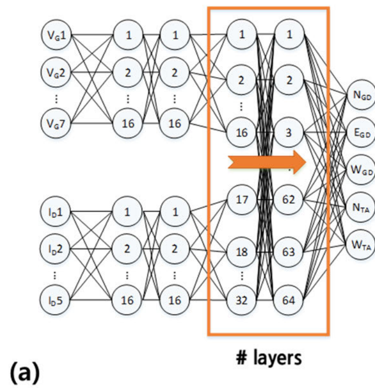
TABLE 2. Comparison of model performance according to the number of layers in the separated part.

Separated part		Merged part		Conformity (%)
# of nodes	# of layers	# of nodes	# of layers	
4	2	32	1	69.1
4	3	32	1	45.9
4	4	32	1	46.0
4	5	32	1	40.9
8	2	32	1	87.0
8	3	32	1	86.8
8	4	32	1	84.6
8	5	32	1	78.4
16	2	32	1	93.0
16	3	32	1	94.0
16	4	32	1	91.3
16	5	32	1	90.5

B. OPTIMIZATION OF MODEL STRUCTURE

The performance of different model structures was compared for current–voltage double-type input data. Models with different numbers of layers and nodes were examined to identify which of the following is more effective—a model with deep layers consisting of a small number of nodes or a model with shallow layers consisting of a large number of nodes. Figure 8 shows the effect of the number of nodes in the separated hidden layer. The depth of the separated and merged hidden layers were 2 and 1, respectively. The accuracy increased as the number of nodes increased. When there were 20,000 training data with ten input nodes and five output nodes, a model complexity of at least 100 nodes was required. Details about the model structure and performance are summarized in Table 1.

We also confirmed the trend depending on the depth of the separated hidden layer with 4, 8, and 16 nodes (Figure 9 and Table 2). The performance is degraded as the depth is increased. Especially, the model with four nodes shows the greatest decrease in performance, as the complexity of the model was too low. In the model with eight or sixteen nodes, the performance improvement by increasing the number of layers was insignificant and tended to decrease slightly. It implies that increasing the number of nodes is more advantageous than staking the separated hidden layer deeper, because information about the correlation between



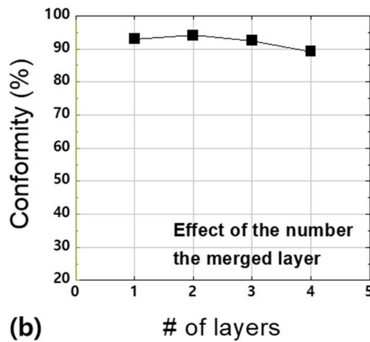
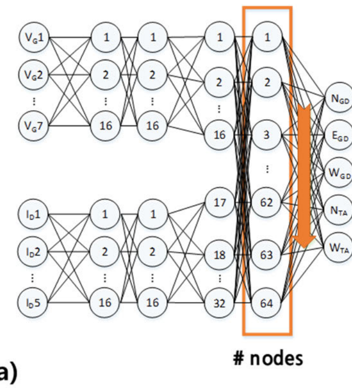


FIGURE 10. Effect of increasing the number of nodes in the last hidden layer from 4 to 512. (a) Model structure increasing width of the merged part and (b) performance.



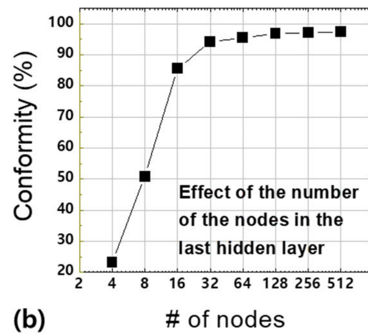


FIGURE 11. Effect of increasing the number of layers in the merged part. Each layer has 32 nodes. (a) Model structure increasing depth of the merged part and (b) performance.

TABLE 3. Comparison of model performance according to the number of nodes in the last hidden layer.

Separated part		Merged part			Conformity (%)
# of nodes	# of layers	# of nodes (layer1)	# of nodes (layer2)	# of layers	
16	2	32	4	2	23.3
16	2	32	8	2	50.8
16	2	32	16	2	85.6
16	2	32	32	2	94.1
16	2	32	64	2	95.4
16	2	32	128	2	96.8
16	2	32	256	2	97.1
16	2	32	512	2	97.4

TABLE 4. Comparison of model performance according to the number of layers in the merged part.

Separated part		Merged part		Conformity (%)
# of nodes	# of layers	# of nodes	# of layers	
16	2	32	1	93.0
16	2	32	2	94.1
16	2	32	3	92.5
16	2	32	4	89.2

I–V characteristics and DOS parameters is lost when the number of nodes is decreased.

With respect to the configuration of the merged hidden layer, it was confirmed that the number of nodes in the last layer had a dominant effect on performance. Figure 10 and Table 3 show the model performance according to the number of nodes in the last hidden layer. Performance improved when the number of nodes was increased, and this improvement plateaued when the number of nodes exceeded 32.

This may be because the regression model outputs quantitative values, unlike the classification model. If the number of nodes is too small, it is difficult to obtain accurate values due to interference between output nodes. Therefore, the number of nodes in the last hidden layer should be maintained above a certain number so that the values in the output layer do not become entangled.

Figure 11 and Table 4 show the performance according to the number of merged layers. The number of nodes in each layer is 32. This trend is similar to the case where the layers are increased in the separated part with 16 nodes in Fig. 9 (black symbol). As the number of nodes is sufficient, increasing the number of layers has little or no impact on the performance, and may even lead to a slight decrease in performance.

Based on this trend, the proposed model uses only 4 layers in total, 2 layers each in the separated part and merged part. The number of nodes in the separated part is 16 for each branch, and the merged part is composed of layers with 32, 256 nodes.

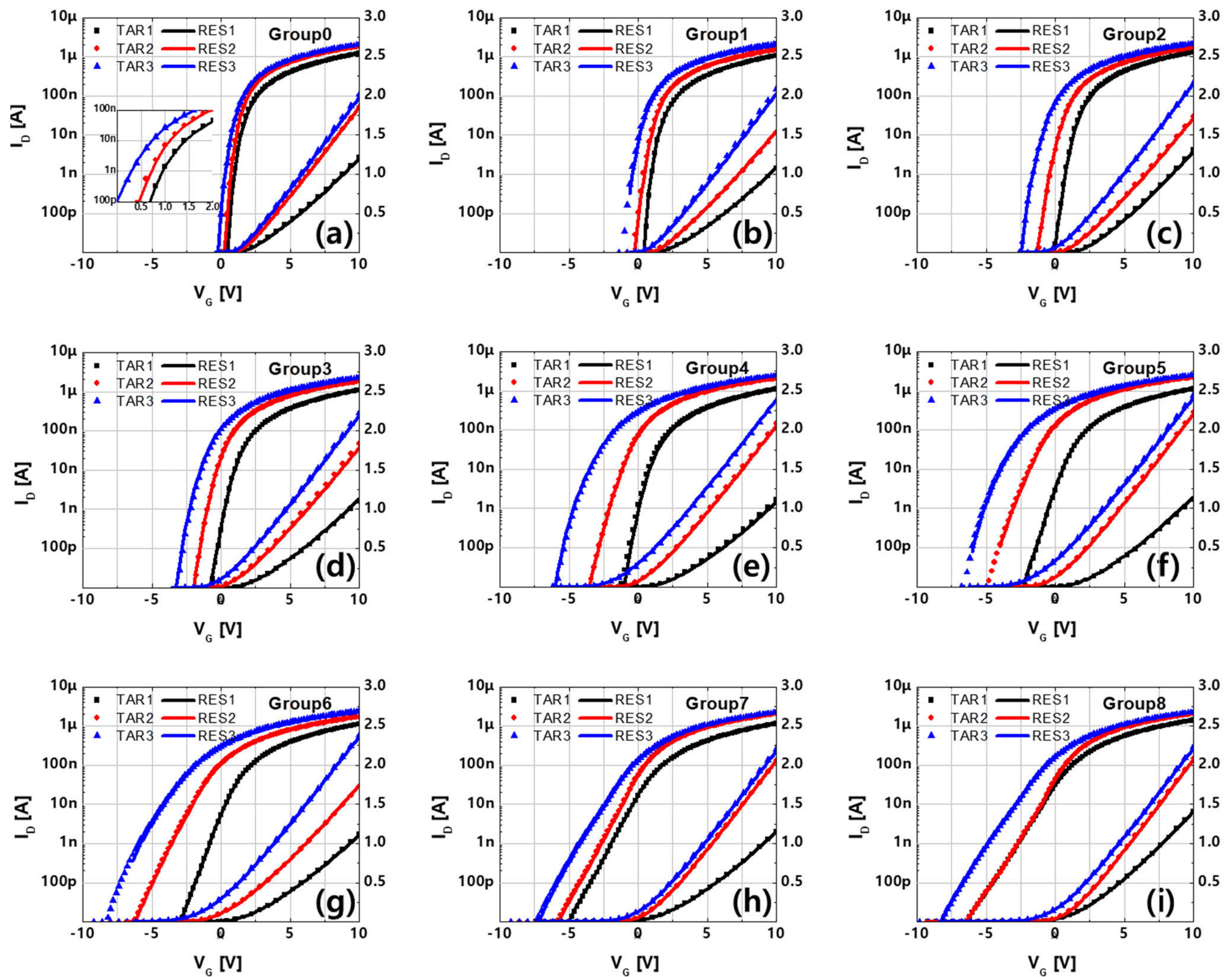


FIGURE 12. I–V characteristics of test set (symbol) and the decoded curve (line) by interpolation using the DOS parameters from the proposed DNN model. Curves of three cases with the largest, middle, and smallest threshold voltages for each group. SS of each group is 0.06–0.3, 0.3–0.4, 0.4–0.5, 0.5–0.6, 0.6–0.8, 0.8–1.0, 1.0–1.4, 1.4–1.8, more than 1.8.

C. CURVE FITTING ACCURACY

The model's purpose is to identify DOS information based on the measured I-V characteristics. The proposed model's performance is evaluated by the difference between a given input I-V curve and a decoded curve based on the model's output DOS parameters. It is important to analyze the performance through a comparison of I-V curves for two reasons. First, in reality, the only information available to users is the measured I-V curve. It seems that the DOS parameters are correct because the DOS parameters entered into the simulation and the resulting curve are known in the training process, but in reality, there is no correct answer. Since the trap state inside a device cannot be measured directly, we can only infer trap information by calibrating the measured curve. Therefore, the accuracy of the model cannot be estimated by DOS parameters themselves, and the decoding process that converts DOS parameters to I-V curves is indispensable. Second, the user can intuitively judge the model performance and select the characteristics to focus on depending on the application. For example, it is important to find an accurate threshold voltage for a device to enable the on/off switching function. Or users need a precise prediction of the on-current level, because the brightness of the pixel in a display is determined by the driving transistor's on-current. The converted I-V curve directly informs the user of device characteristics such as threshold voltage, SS, and on-current. The user can modulate the loss function by adjusting the joint MSE weight of the DOS parameters and compare the model's performance through the I-V curve. The accuracy of V_{TH} can be improved by increasing the loss proportion of donor-like Gaussian states parameters N_{GD}, E_{GD}, and W_{GD}. An effective way to reduce the on-current difference is to increase the weight of N_{TA} and W_{TA}.

A threshold voltage difference under 0.2 V and on-current error under 5% are allowed for calculating conformity. Figure 12 shows the I–V characteristics of the test set and decoded curves by interpolation using DOS parameters from the proposed DNN model. The 14400 test curves are divided

into nine groups by SS. Three curves with maximum, median, and minimum V_{TH} among the curves in each group are depicted. The decoded curves calibrate the test curves with large V_{TH}, SS, and on-current deviations well. Using this model, the DOS parameters can be extracted quickly and easily.

IV. CONCLUSION

In this paper, we proposed a DNN model that automatically extracts the subgap state in the channel of an oxide TFT based on the I–V characteristic. We used 14400 supervised learning data obtained by TCAD simulation to train the model to correlate DOS parameters and I–V characteristics. The proposed model can determine the subgap state distribution of samples with a wide range of deviation of approximately 8 V in the threshold voltage and one order of magnitude in the on-current. Users with no background knowledge can analyze measured I-V characteristics and extract DOS parameters for quick and easy curve calibration.

In addition, we proposed an interpolation technique that converts a DOS set from the DNN model into an I–V curve without TCAD simulation. The interpolation distance was defined by the difference between a result DOS and its 32 nearest DOS combinations, and the decoding curve was interpolated using the 32 corresponding curves. After interpolation, the decoding curve was compared with a target curve to evaluate the curve fitting accuracy. More than 97% of all the samples showed a threshold voltage and on-current error within 0.2 V and 5%, respectively. Automated DOS extraction using the proposed model was more than 10 million times faster than manual extraction using repetitive TCAD simulation. On this basis, we verified that the proposed model has the accuracy and efficiency required to replace the TCAD curve fitting process, thereby enabling the automatic analysis of many devices.

This paper included two practical insights regarding data preprocessing and model structure of the regression models. First, input data of the current–voltage double type is more advantageous than the commonly used I–V curve image or log-scale current value. The log current data includes noisy off current and the on and off regions are not divided on the basis of a specific value of V_G that is common for all devices because the V_{TH} of the dataset ranges from –7 V to 1 V. Unlike log current-type input, the voltage type does not deteriorate owing to noise current or noise filtering. The voltage type data has valid values regardless of the device's V_{TH}. The difference lies only in whether the gate voltage sweep range in the subthreshold region is widened or narrowed. And current values in the linear scale reflect a difference between curves in the on-region more efficiently than using log scale currents. The proposed double-type input data takes advantage of both linear current and voltage type information, and as a result, model performance can be improved through appropriate data preprocessing.

Second, increasing the number of nodes in the DNN model is advantageous, although the depth of the hidden layer is reduced. Unlike classification models, regression models should be able to fit quantitative values accurately. When the number of nodes is small, the model cannot adjust the output value finely, and the trend of the entire weight may be lumped. Increasing the number of nodes in the last hidden layer helps improve performance. In the proposed model, each output node estimates the values of five parameters that determine the DOS distribution. In a regression model with more than two output nodes, a last hidden layer with a small number of nodes results in interference between the output information. Therefore, increasing the number of final hidden layers can improve the performance of the regression model in fitting the target data for each node accurately.

REFERENCES

[1] S. J. Kim, S. Yoon, and H. J. Kim, "Review of solution-processed oxide thin-film transistors," *Jpn. J. Appl. Phys.*, vol. 53, no. 2S, Feb. 2014, Art. no. 02BA02.

[2] A. Kumar, A. K. Goyal, and N. Gupta, "Thin-film transistors (TFTs) for highly sensitive biosensing applications: A review," *ECS J. Solid State Sci. Technol.*, vol. 9, no. 11, 2020, Art. no. 115022.

[3] J. W. Park, B. H. Kang, and H. J. Kim, "A review of low-temperature solution-processed metal oxide thin-film transistors for flexible electronics," *Adv. Funct. Mater.*, vol. 30, no. 20, 2020, Art. no. 1904632.

[4] L. Zhang, H. Yu, W. Xiao, C. Liu, J. Chen, M. Guo, H. Gao, B. Liu, and W. Wu, "Strategies for applications of oxide-based thin film transistors," *Electronics*, vol. 11, no. 6, p. 960, Mar. 2022.

[5] K. Nomura, H. Ohta, A. Takagi, T. Kamiya, M. Hirano, and H. Hosono, "Room-temperature fabrication of transparent flexible thin-film transistors using amorphous oxide semiconductors," *Nature*, vol. 432, no. 4016, pp. 488–492, Nov. 2004.

[6] P. Heremans, A. K. Tripathi, A. de Jamblinne de Meux, E. C. Smits, B. Hou, G. Pourtois, and G. H. Gelinck, "Mechanical and electronic properties of thin-film transistors on plastic, and their integration in flexible electronic applications," *Adv. Mater.*, vol. 28, no. 22, pp. 4266–4282, 2016.

[7] K. Ide, K. Nomura, H. Hosono, and T. Kamiya, "Electronic defects in amorphous oxide semiconductors: A review," *Phys. Status Solidi (A)*, vol. 216, no. 5, Mar. 2019, Art. no. 1800372.

[8] M. M. Billah, M. M. Hasan, M. Chun, and J. Jang, "TCAD simulation of dual-gate a-IGZO TFTs with source and drain offsets," *IEEE Electron Device Lett.*, vol. 37, no. 11, pp. 1442–1445, Nov. 2016.

[9] S. K. Dargar and V. M. Srivastava, "Design and analysis of IGZO thin film transistor for AMOLED pixel circuit using double-gate tri active layer channel," *Heliyon*, vol. 5, no. 4, Apr. 2019, Art. no. e01452.

[10] Y. Kim, M. Bae, W. Kim, D. Kong, H. K. Jung, H. Kim, and D. H. Kim, "Amorphous InGaZnO thin-film transistors—Part I: Complete extraction of density of states over the full subband-gap energy range," *IEEE Trans. Electron Devices*, vol. 59, no. 10, pp. 2689–2698, Oct. 2012.

[11] H. Oh, K. Cho, and S. Kim, "Effect of physical densification on subgap density of states in amorphous InGaZnO thin films," *Superlattices Microstruct.*, vol. 121, pp. 33–37, Sep. 2018.

[12] J. T. Jang, D. Ko, S. Choi, H. Kang, J.-Y. Kim, H. R. Yu, G. Ahn, H. Jung, J. Rhee, H. Lee, S.-J. Choi, D. M. Kim, and D. H. Kim, "Effects of structure and oxygen flow rate on the photo-response of amorphous IGZO-based photodetector devices," *Solid-State Electron.*, vol. 140, pp. 115–121, Feb. 2018.

[13] J. T. Jang, J. Park, B. Du Ahn, D. M. Kim, S.-J. Choi, H.-S. Kim, and D. H. Kim, "Effect of direct current sputtering power on the behavior of amorphous indium-gallium-zinc-oxide thin-film transistors under negative bias illumination stress: A combination of experimental analyses and device simulation," *Appl. Phys. Lett.*, vol. 106, no. 12, 2015, Art. no. 123505.

[14] M. Mativenga, F. Haque, M. M. Billah, and J. G. Um, "Origin of light instability in amorphous IGZO thin-film transistors and its suppression," *Sci. Rep.*, vol. 11, no. 1, pp. 1–12, 2021.

[15] J. G. Um, M. Mativenga, and J. Jang, "Mechanism of positive bias stress-assisted recovery in amorphous-indium-gallium-zinc-oxide thin-film transistors from negative bias under illumination stress," *Appl. Phys. Lett.*, vol. 103, no. 3, Jul. 2013, Art. no. 033501.

[16] J. Sheng, J. Park, D. W. Choi, J. Lim, and J. S. Park, "A study on the electrical properties of atomic layer deposition grown InO_x on flexible substrates with respect to N_2O plasma treatment and the associated thin-film transistor behavior under repetitive mechanical stress," *ACS Appl. Mater. Interfaces*, vol. 8, no. 45, pp. 31136–31143, 2016.

[17] Y. Seo, H.-S. Jeong, H.-Y. Jeong, S. Park, J. T. Jang, S. Choi, D. M. Kim, S.-J. Choi, X. Jin, H.-I. Kwon, and D. H. Kim, "Effect of simultaneous mechanical and electrical stress on the electrical performance of flexible In-Ga-Zn-O thin-film transistors," *Materials*, vol. 12, no. 19, p. 3248, Oct. 2019.

[18] M. Hasan, M. M. Billah, and J. Jang, "Tensile stress effect on performance of a-IGZO TFTs with source/drain offsets," *IEEE Electron Device Lett.*, vol. 39, no. 2, pp. 204–207, Feb. 2018.

[19] M. M. Billah, M. M. Hasan, and J. Jang, "Effect of tensile and compressive bending stress on electrical performance of flexible a-IGZO TFTs," *IEEE Electron Device Lett.*, vol. 38, no. 7, pp. 890–893, Jul. 2017.

[20] M. Kimura, T. Nakanishi, K. Nomura, T. Kamiya, and H. Hosono, "Trap densities in amorphous-$InGaZnO_4$ thin-film transistors," *Appl. Phys. Lett.*, vol. 92, no. 13, 2008, Art. no. 133512.

[21] K. A. Stewart, V. Gouliouk, J. M. McGlone, and J. F. Wager, "Side-by-side comparison of single- and dual-active layer oxide TFTs: Experiment and TCAD simulation," *IEEE Trans. Electron Devices*, vol. 64, no. 10, pp. 4131–4136, Oct. 2017.

[22] Z.-W. Shang, Q. Xu, G.-Y. He, Z.-W. Zheng, and C.-H. Cheng, "Effect of plasma oxidation on tin-oxide active layer for thin-film transistor applications," *J. Mater. Sci.*, vol. 56, no. 10, pp. 6286–6291, Apr. 2021.

[23] B.-S. Kim, H.-J. Jeong, K.-L. Han, W.-B. Lee, Y.-S. Kim, and J.-S. Park, "Soft recovery process of mechanically degraded flexible a-IGZO TFTs with various rolling stresses and defect simulation using TCAD simulation," *IEEE Trans. Electron Devices*, vol. 67, no. 2, pp. 535–541, Feb. 2020.

[24] Sivaco. *Atlas User's Manual Device Simulation Software*. Accessed: Aug. 6, 2016. [Online]. Available: https://www.silvaco.com

[25] P. B. L. Meijer, "Neural network applications in device and subcircuit modelling for circuit simulation," in *Philips Electronics*. Eindhoven, The Netherlands: Philips Research Laboratories, 1996.

[26] L. Zhang and M. Chan, "Artificial neural network design for compact modeling of generic transistors," *J. Comput. Electron.*, vol. 16, no. 3, pp. 825–832, 2017.

[27] M. Li, O. Irsoy, C. Cardie, and H. G. Xing, "Physics-inspired neural networks for efficient device compact modeling," *IEEE J. Explor. Solid-State Comput. Devices Circuits*, vol. 2, pp. 44–49, 2016.

[28] K. Lamamra and S. Berrah, "Modeling of MOSFET transistor by MLP neural networks," presented at the Int. Conf. Electr. Eng. Control Appl. Cham, Switzerland: Springer, Nov. 2016, pp. 407–415.

[29] F. Klemme, J. Prinz, V. M. van Santen, and J. Henkel, and H. Amrouch, "Modeling emerging technologies using machine learning: Challenges and opportunities," in *Proc. 39th Int. Conf. Comput.-Aided Design*, 2020, pp. 1–9.

[30] 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. Simulation Semiconductor Processes Devices (SISPAD)*, Sep. 2019, pp. 1–4.

[31] H. M. Kocak, A. T. Naskali, O. Pinarer, and J. Mitard, "Detecting transistor defects in medical systems using a multi model ensemble of convolutional neural networks," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2021, pp. 4731–4737.

[32] J. Kim, S. J. Kim, J.-W. Han, and M. Meyyappan, "Machine learning approach for prediction of point defect effect in FinFET," *IEEE Trans. Device Mater. Rel.*, vol. 21, no. 2, pp. 252–257, Jun. 2021.

[33] M.-H. Oh, M.-W. Kwon, K. Park, and B.-G. Park, "Sensitivity analysis based on neural network for optimizing device characteristics," *IEEE Electron Device Lett.*, vol. 41, no. 10, pp. 1548–1551, Oct. 2020.

[34] H. Yun, J.-S. Yoon, J. Jeong, S. Lee, H.-C. Choi, and R.-H. Baek, "Neural network based design optimization of 14-nm node fully-depleted SOI FET for SoC and 3DIC applications," in *Proc. 4th IEEE Electron Devices Technol. Manuf. Conf. (EDTM)*, Apr. 2020, pp. 1–4.

[35] T.-L. Wu and S. B. Kutub, "Machine learning-based statistical approach to analyze process dependencies on threshold voltage in recessed gate AlGaN/GaN MIS-HEMTs," *IEEE Trans. Electron Devices*, vol. 67, no. 12, pp. 5448–5453, Dec. 2020.

YUNYEONG CHOI (Member, IEEE) was born in Seoul, South Korea, in 1992. She received the B.S. degree in electronic electrical engineering from Ewha Womans University, in 2017, where she is currently pursuing the Ph.D. degree. She researched the model and simulation for the silicon-based TFT, oxide TFT, and OLED. She had involved in the industrial-academic projects about OLED display panel with LG display, from 2016 to 2019. She has conducted modeling and simulation of flexible TFT under mechanical stress and the physically unclonable function hardware chip based on ReRAM. Her current research interest includes deep learning model for analysis transistor characteristics.

WOOKYUNG SUN received the B.S., M.S., and Ph.D. degrees in electronics engineering from Ewha Womans University, Seoul, South Korea, in 1999, 2001, and 2014, respectively. From 2001 to 2009, she was with HYNIX Semiconductor Inc., South Korea. She is currently a Visiting Professor with Seoul National University. Her current research interests include the capless DRAM and memristor devices for PUF or neuromorphic engineering.

JISUN PARK received the B.S., M.S., and Ph.D. degrees in electronics engineering from Ewha Womans University, Seoul, South Korea, in 1999, 2001, and 2005, respectively. She started as a Process and Device Engineer with SK Hynix Inc., in 2005, where she worked on the development of the world's first 60nm level DRAM and various mobile application memory. From 2010 to 2013, she was a Senior Research Engineer with Synopsys and also as a Research Fellow with Kookmin University, from 2013 to 2017. Since 2017, she has been a Research Fellow Member of the Department of Electronic and Electrical Engineering, Ewha Womans University. Her current research interests include the modeling and simulation about unit device and circuits for high density memory and display application.

HYUNGSOON SHIN (Senior Member, IEEE) was born in Seoul, South Korea, in 1959. He received the B.S. degree in electronics engineering from Seoul National University, in 1982, and the M.S. and Ph.D. degrees in electrical engineering from The University of Texas at Austin, in 1984 and 1990, respectively. From 1990 to 1994, he was with LG Semiconductor Company Ltd., South Korea, where he worked on the development of 64M DRAM, 256M DRAM, 4M SRAM, and 4M FLASH memory. Since 1995, he has been a Faculty Member of the Department of Electronic and Electrical Engineering, Ewha Womans University, Seoul. His current research interests include new processes, devices, and circuit developments and modeling based on Si, both for high density memory and RF IC. He has published numerous journal articles on implant profile models, mobility models, deep-submicron MOSFET structure analysis, current crowding effect in diagonal MOSFET, hot-carrier degradation, alpha-particle-induced soft error, and MRAM. He is a Senior Member of the Institute of Electronics Engineers of Korea. He received the Technical Excellence Award from the Semiconductor Research Corporation (SRC), in 1991.