Mapping Oxidation and Wafer Cleaning to Device Characteristics Using Physics-Assisted Machine Learning

Sparsh Pratik, Po-Ning Liu, Jun Ota, Yen-Liang Tu, Guan-Wen Lai, Ya-Wen Ho, Zheng-Kai Yang, Tejender Singh Rawat, and Albert S. Lin*

ABSTRACT: It is always highly desired to have a well-defined relationship between the chemistry in semiconductor processing and the device characteristics. With the shrinkage of technology nodes in the semiconductors roadmap, it becomes more complicated to understand the relation between the device electrical characteristics and the process parameters such as oxidation and wafer cleaning procedures. In this work, we use a novel machine learning approach, i.e., physics-assisted multitask and transfer learning, to construct a relationship between the process conditions and the device capacitance voltage curves. While conventional semiconductor processes and device modeling are based on a physical model, recently, the machine learning-based approach or hybrid approaches have drawn significant attention. In general, a huge amount of data is required to train a machine learning model. Since producing data in the semiconductor industry is not an easy task, physics-assisted artificial intelligence has become an obvious choice to resolve these issues. The predicted C−V uses the hybridization of physics, and machine learning provides improvement while the coefficient of determination (R²) is 0.9442 for semisupervised multitask learning (SS-MTL) and 0.9253 for transfer learning (TL), referenced to 0.6108 in the pure machine learning model using multilayer perceptrons. The machine learning architecture used in this work is capable of handling data sparsity and promotes the usage of advanced algorithms to model the relationship between complex chemical reactions in semiconductor manufacturing and actual device characteristics. The code is available at https://github.com/albertlin11/moscapssmtl.

1. INTRODUCTION

Predicting device characteristics as the function of semiconductor process conditions is a difficult but important task. The difficulty lies in the fact that there are many chemical reactions and multidisciplinary knowledge is necessary in semiconductor manufacturing procedures. To use the analytical approaches in chemistry and physics to tackle the modeling of semiconductor process and its relation to the device characteristics, the results are always away from the experiments. In some cases, the model is even absent for this purpose. With the rise of machine learning, it becomes more feasible to use machine learning algorithms to describe the complex relations between semiconductor processes, such as etching, deposition, oxidation, wafer cleaning, and lithography, and semiconductor device characteristics.

Metal oxide semiconductor field-effect transistors (MOSFETs) play a vital role in today’s integrated circuit (IC) technologies. Ideally, for a MOS device, it is desired that the electrical characteristics should be controlled by the bulk, not by the interfaces. By analyzing the undesirable charges present in the interface, capacitance–voltage (C−V) characteristics help in determining the stability and overall performance. It
has been seen that the silicon and silicon dioxide interface is badly affected by the interface trap and fixed charge. Thus, it becomes of utmost importance to study the process parameters, which are in turn influencing the electrical characteristics of devices. Industry 4.0 gives an idea about smart manufacturing in which foundries are looking to artificial intelligence and the Internet of Things for reducing the cost and maintaining state-of-the-art fabrication quality. Depending on the different capacitance present in MOS capacitors (MOSCAPs), $C-V$ is divided into three major sections: accumulation, depletion, and inversion. The total capacitance per area ($C_T$) is expressed as the series combination of surface differential capacitance per unit area ($C_s$) and oxide capacitance per unit area ($C_{ox}$). As transistors shrink and the process becomes more complex, there are more process steps in semiconductor manufacturing. Conventionally, people use theory-based approaches to describe device behaviors, but mapping the device characteristics to real process parameters such as oxidation, RCA cleaning, etc., can be difficult without the help of regression or machine learning (ML). With the rise of ML in electronic design automation (EDA) in recent years, it becomes more practical to incorporate the process parameters, such as annealing temperature and plasma power, into the compact model and circuit simulation.

ML models have been working excellently in many commercial business applications like fraud detection, machine translation, and recommendation systems. The drawback of ML-based compact models (CMs) is that a huge set of labeled data is required to implement these applications. Due to the requirement of a large set of labeled data for ML models and the complexity that occurs in pure physical modeling, it may not be feasible to design either a pure ML or pure physical device model for emerging technology nodes with process-aware capability. The best modeling technique for the semiconductor industry is to mix the idea of ideal physics with ML algorithms, and these limited practical labeled data available. This type of hybrid model has been accomplished in many diverse fields like in the chemical synthesis of pharmaceutical products, water management systems, and the semiconductor design automation industry. Specific to IC foundries, the availability of noise-free practical silicon data is a matter of great concern. A semiconductor MOS device requires a large set of fabrication machines which costs a lot of time and money. With the limited resource of the data, the physics-based ML models recognize the pattern and behavior of the physical processes; ergo, they lead to state-of-the-art modeling. The device physics can be processed with the hyperparameter of ML-like calibration of the current–voltage characteristics of a tunnel FET with the help of different activation functions. This paradigm shift in semiconductor modeling gives a better convergence and improves the accuracy of the model. The neural network simulation models can also be extended to the entire family range, like a unified resistive random access memory model for the whole memristor family.

The efficacy of supervised learning is very high but comes with the requirement of a large amount of label data. Semisupervised multitask learning (SS-MTL) and transfer learning (TL) are advanced ways to look into this problem. For a classification-type problem, these algorithms reached state-of-the-art performance. In the semiconductor industry, TL-based image defect classification is quite common in wafer
manufacturing industries.29 As the performance of models has achieved the human-level classification, this gives a boost to the manufacturer in lowering the cost and increasing yield. The performance of the TL algorithm on regression problems is also significant, e.g. predicting the phonon characteristics of a wideband semiconductor using its electronic properties.30 It is also successfully implemented in determining the formation energy of the material, further improving material properties.31 With the fewer data available, the study of some new materials can also be analyzed using TL.32,33 C. Park et al. presented the application of multitask learning (MTL) in determining the quality of wafers coming from different process chambers.34 Regardless of the different process parameters in various chambers, the operation is the same. Thus, the idea of MTL reduces the necessity of separate models for each process variation35 and helps in the production of semiconductor components.

We are presenting a physics-based stimulated ML regression model that will ease out the data dependency. One model is related to the sharing layer-based semisupervised multitask learning approach,36−38 and another is the TL-based approach.39,40 In this research paper, we present an ML-based model for different tasks. Here tasks are categorized, and they are based on the prediction of $C−V$ for intermediate frequencies. The measured data will be exclusively used for testing and not for training. This implies that our presented model is only trained on the measured data at low and high frequencies. The physical modeling for all frequencies will supplement the ML model to improve the prediction accuracy. The knowledge of the physical model will either be shared or transferred to facilitate prediction, depending on either SS-MTL or TL is used. Our presented models outclass the baseline multilayer perceptron (MLP), which is the most famous ML algorithm in the electronic design automation industry. We have incorporated the effect of dry/wet oxidation, different types of wafer cleaning methods, and different metal deposition techniques, and therefore our compact modeling is process-aware. With a little more computational power, our models can be easily expanded to any large number of process parameters in semiconductor manufacturing. The initial results of this work can be found in our previous conference presentation38 and student theses.41,42

2. METHODOLOGY

A three-layered metal-oxide-semiconductor structure is fabricated with different process parameters. We have considered different types of wafer cleaning, thermal oxidation, metal deposition, and metal electrode lithography as process parameters. The entire flow from fabrication and measurement to machine learning is illustrated in Figure 1.

2.1. Data Collection. We have 144 288 theoretical data and 4896 measurement data. The input variables are the frequency, area, $\text{RCA/standard clean}$, dry/wet oxidation, $\text{sputter/e-gun Al}$, and voltage. The frequencies are 1, 3, 5, 10, 50, or 100 kHz. The area is from 0.005, 0.0049, 0.01, 0.0102, 0.0025, or 0.0026 mm$^2$. The applied voltage across the MOSCAP is from $−4$ to 2.5 V. The training data are 1 and 100 kHz, and the test data is 3, 5, 10, and 50 kHz. Out of the training data, 0.01 mm$^2$ data served as the validation set. The output variable is capacitance in units of picofarad.

2.2. Sample Fabrication. The reliability and overall yield strongly depend on the amount of contamination present on the wafer. Consequently, it becomes important to study the effect of cleaning. First, a p-type wafer is cleaned in two different types of processes: one is $\text{RCA cleaning}$,43 and another is $\text{standard cleaning}$. Chemicals, i.e., $\text{NH}_4\text{OH}$, $\text{H}_2\text{O}_2$, $\text{HCl}$, $\text{H}_2\text{SO}_4$, etc., are used in different concentrations in these two cleaning methods. In the standard cleaning, first, a composition of $\text{NH}_4\text{OH}$, $\text{H}_2\text{O}_2$, and $\text{HCl}$ in the proportion of 1:4:20 is used for removing particles and organic contamination.
chemical treatments, there may be a chance of native oxide growth. To eliminate this residue, the wafers are dipped into a 1:50 proportionate solution of HF and H₂O. After every chemical treatment of wafers, a quick dump rinser is also used for maintaining wafer quality. The entire wafer’s cleaning process is carried out on a wet bench followed by a spin dryer. Once the cleaning is done, the next step is to grow oxide. There are many oxide deposition methods like thermal oxidation, plasma anodization, and vapor phase reaction. Among these methods, thermal oxidation is a widely used process in industries. This wet or dry oxidation deposition of 200 Å thickness is completed through a horizontal diffusion furnace. The top metal layer of aluminum is deposited through two different types of physical vapor deposition (PVD) techniques. One is the AST E-GUN PEVA-600I electron-beam PVD (E-gun), and another is the FSE Cluster PVD sputter.

The deposition in the E-gun takes place in the presence of an electric field where high-energy electron beams are generated through thermionic emission. These accelerated and focused beams fall on the source material that is kept in a crucible. The source material is evaporated and deposited on the substrate. In sputtering, the high-energy plasma hits the target material, and after condensation, a thin layer of the target is get deposited on the substrate. After the deposition of the MOS structure, we have done the lithography of all the samples. The square and circle shapes of electrodes have been used. During lithography, a thin positive photoresist layer of thickness 8500 ± 150 Å is deposited by using TEL CLEAN TRACK MK-8. An exposure dose of 2000 J/m² is used in the Canon PPA-3000iS+ 1-Line stepper. It has a resolution of nearly 0.35 μm. Again, the track machine is used for the removal of the exposed photoresist. Once the lithography is done, the electrodes are formed by dry etching, and ozone ash is used afterward for residual resist removal. TCP 9600SE etching machine gives the desired structure. Due to the limitation of lithography, the wafers are sliced into a small die area of 2 cm × 2 cm. During the frontend process, native oxide gets developed on the backside of the samples. To handle such a scenario, an etching machine is used again, and a thin layer of the aluminum film is deposited by the AST PEVA 600I e-gun. The process flow of sample fabrication is depicted in Figure 2 and can further be seen from the TEM image, as shown in Figure 3.

2.3. C–V Measurement. We use an Agilent 4284A for MOSCAP measurement. One probe is placed on the top electrode, which will give the AC signal, and the back electrode is connected by a chuck and grounded for measurement. Afterward, the semiconductor device analyzer software (EasyEXPERT) is used to adjust the sweep voltage range (from −4 to 2.5 V, step voltage 0.13 V), integration time (medium), and measurement frequency (from 1 kHz to 1 MHz). Our sample is measured under illumination using a Schott ACE light source EKE 150W halogen lamp. A detailed description of fabrication and measurement can be found in student theses.41,42

2.4. Theoretical C–V Generation. Typical C–V characteristics of a MOS capacitor are affected by many physical parameters.8,9 Typically, electrons (n) and holes (p) as mobile charges and ionized acceptor (N⁺) and donor (N⁻) as fixed charges are present in the silicon. Poisson’s equation can be written as:

\[
\nabla \cdot \left( \varepsilon \nabla V \right) = \rho_f \tag{9}
\]
\[
\frac{d^2 \psi}{dx^2} = -\frac{q}{\varepsilon_s} [p(x) - n(x) + N_a^+(x) - N_a^-(x)]
\]

\[
= -\frac{q}{\varepsilon_s} [N_s(e^{-q\psi/kT} - 1) - \frac{n_i^2}{N_s}(e^{q\psi/kT} - 1)]
\]

Equations 1 and 2 describe the profile of the electric field and are represented by \( \xi(x) = \left( \frac{d \psi}{dx} \right)^2 \)

\[
\frac{kTN}{e} \left[ \left( e^{-q\psi/kT} + \frac{q\psi}{kT} - 1 \right) + \frac{n_i^2}{N_s}(e^{q\psi/kT} - \frac{q\psi}{kT} - 1) \right]
\]

According to Gauss’s law, the total charge per unit area \( (Q_x) \) induced in the semiconductor is defined by 4. At the limiting condition of \( x = 0 \), \( \psi = \psi_0 \) and \( \xi = \xi_0 \), the total charge density will be the surface charge density, and potential will be the surface potential: 9

\[
Q_x = -\varepsilon_s \xi_x
\]

\[
= \pm \sqrt{2e_kTN} \left[ e^{-q\psi/kT} + \frac{q\psi}{kT} - 1 \right]^{1/2}
\]

The overall capacitance can be obtained by

\[
\frac{1}{C} = \frac{1}{C_{ox}} + \frac{d(\psi)}{d(-Q_x)} = \frac{1}{C_{ox}} + \frac{1}{C_{si}}
\]

The gate voltage consists of the voltage drop across the oxide and the surface potential. In the ideal case, flat-band voltage \( (V_{fb}) \) is zero, but practically it depends on metal work function and traps. A semiempirical voltage shift in flat band voltage calculation provides a better closeness with silicon data C–V. In this work, by comparing to the training set data, an empirical voltage shift of \(-0.78 \text{ V} \) is added to \( V_{fb} \). The voltage across MOSCAP can be expressed as

\[
V - V_{fb} = V_{ox} + \psi_s = \frac{-Q_x}{C_{ox}} + \psi_s
\]

When the gate voltage is positive, the inversion charges start to increase and slowly enter into the inversion region. The inversion region is dominated by minority carriers. Hence, 4 can be simplified and shown by 7:

\[
Q_{inv} = \sqrt{2e_kTN} \left[ \frac{n_i^2}{N_s}(e^{q\psi/kT} - \frac{q\psi}{kT} - 1) \right]^{1/2}
\]

According to our experiment, we considered gate voltage from 0 to 2.5 \( \text{V} \), and the lowest frequency as 1 kHz. For a specific frequency, if the gate voltage is increased, we see an increment in the inversion charge. On further increase in the voltage, inversion charges become saturated, and the threshold of \( \Delta Q_{inv} \) is reached. This is the main reason behind the flat capacitance curve in the inversion region. Since the measurement is done in the presence of light, there will be some movement in the charge carrier, and it is expressed as

\[
\Delta Q_{inv\theta} = \Delta Q_{inv\theta}(1 - \exp(k \times t))
\]

where \( \Delta Q_{inv\theta} \) is the threshold of \( \Delta Q_{inv} \) for 1 kHz, \( \Delta Q_{inv\theta} \) is the threshold for a specific frequency \( (f) \), \( k \) is the constant, and \( t \) is the transfer time which is equal to the period \( T = 1/f \). In our measurement, 1 kHz of the C–V characteristic behaves like the low-frequency characteristics under illumination. In calculation, we assume 100 kHz C–V is the lowest frequency that will show ideal high-frequency C–V, and this determines the \( \Delta Q_{inv} / \Delta Q_{inv\theta} \) at 1 kHz, which in turn gives the value of \( k \) as the basis of a low-frequency signal. Transferred carriers establish an exponential relationship with the transfer time.

The threshold value of \( \Delta Q_{inv\theta} \) for different frequencies is calculated using 8.

When \( \Delta Q_{inv} \) reaches \( \Delta Q_{inv\theta} \), the capacitance value no longer increases due to the fact that the lateral transport of the minority electrons from the illuminated region surrounding the metal electrode becomes insufficient to sustain the \( \Delta Q_{inv} \). We can denote the corresponding voltages as \( V_{inv\theta} \). When the gate voltage is equal to or larger than \( V_{inv\theta} \) the overall capacitance reaches a fixed value, and this can be observed in Figure 4. The \( R^2 \) between the experiment and theoretical data is 0.9176 in this work. It should be emphasized that the purpose of this work is to show that analytical principles can assist ML, but in the end, ML models, instead of physical models, should be the choice to be used in semiconductor manufacturing. The reason is that, until now, there have not been reasonable models for many manufacturing processes. For example, the relation between wafer cleaning to device C–V is not well described by any physical or chemistry models.
2.5. Machine Learning Algorithm. MLP is the most commonly used algorithm in the electronic area. This model works excellent for a single task, but the problem associated with this is its extension for newer data. Advanced algorithms give leverage to extend the prediction capability for unseen data. The success of an ML lies in the test data performance. Thereby, the concept of overfitting becomes of utmost importance to deal with. The model hyperparameters are listed in Table 2. The first important hyperparameter is the number of hidden layers and the number of neurons present in each layer are considered as two of the most important model hyperparameters. Typically, in a neuron, the incoming signals get multiplied with their respective weight and added with their bias.

\[ y = f_{\text{activation}} \left( \sum_{i=1}^{n} x_i w_i + b \right) \]  

(9)

In (9), \( y \) is the output from the neuron, \( i \) is the total number of features, \( x_i \) is the input, \( W_i \) is the weight, and \( b \) is the bias. Different activation functions introduce nonlinearity in the ML model.

2.5.2. Semisupervised Multitask Learning (SSMTL). The idea behind multitask learning (MTL) is to train the model with similar types of tasks rather than training separate models for each task. The sharing of weights and biases among the models elevates the model performance. Additionally, getting labeled data for every task is very difficult and expensive. If the sizes of the labeled data for the tasks are different, then the methodology is essentially a semisupervised (SS) MTL. Semisupervised learning gives a boost to the model in which unlabeled and labeled data are mixed to enhance the performance of the model. Figure 5b shows the architecture of SSMTL in which two similar types of tasks, i.e., theoretical and measured \( C-V \) characteristics of MOS capacitors, are trained together. The knowledge of theoretical \( C-V \) assists in predicting the values of capacitance at intermediate frequencies.

2.5.3. Transfer Learning. The intuition behind transfer learning is to utilize the knowledge of a previously trained model, as illustrated in Figure 5c, i.e., reusability of a pretrained model.\(^{46}\) For example, the knowledge of riding a bicycle can be used in learning motorcycles. Similarly, in ML-based compact device modeling, the knowledge of training theoretical \( C-V \) can be used in predicting real silicon \( C-V \) data. In transfer learning, the optimized weights and biases of the corresponding pretrained model are transferred to a new model. Hence, the new model constitutes previously trained layers and only needs one or two more layers for any new task. This also helps in saving a lot of computational power.

3. RESULTS AND DISCUSSION

A threshold voltage of any MOS device is the most important parameter to evaluate its operating condition, and it is observed as the onset point of strong inversion. In our experiments, the illumination is the source of minority charge carriers for MOSCAP to achieve the strong inversion, and the capacitance depends on the frequency of that sweeping AC voltage. The illumination leads to optical generation at the metal top electrode to enable the capacitor to go into the inversion region. At the higher frequency, the time period of the AC sweeping voltage signal will be much smaller than the charge lateral transfer time, and thus, the inversion capacitance becomes smaller. Eventually, it becomes crucial to

![Figure 4: Theoretical MOS capacitor C–V characteristics for the frequencies from 1 kHz to 100 kHz.](https://doi.org/10.1021/acsomega.1c05552)

Table 2. List of Model Parameters and Hyperparameters

| Parameter                    | MLP (baseline) | SSMTL | pretrained model | main model |
|------------------------------|----------------|-------|------------------|------------|
| number of hidden layers      | 2              | 2     | 2                | 2\(^a\)     |
| number of neurons (sweep)    | 20–60          | 20–60 | 50               | 20–60      |
| activation function          | Relu           | Relu  | Relu             | Relu       |
| batch size                   | 50             | 50    | 50               | 50         |
| size of training sample      | (4080, 6)      | (124320, 6) | (120240, 6) | (4080, 6) |
| size of test sample          | (9792, 6)      | (9792, 6) | (9792, 6)      | (9792, 6) |
| patience in early stop cycle | 50             | 50    | 50               | 50         |
| number of epoch cycle        | 164            | 179   | 292              | 30         |

\(^{a}\)The first hidden layer is from the pretrained model (number of neurons = 50).
define the C–V as a function of frequency. Our presented model can easily be used for analyzing the dependency of C–V on different frequencies, and simultaneously, it is used for optimizing the fabrication conditions. Generally, a large volume of data is required for the standard deep learning algorithm. Through this paper, we have demonstrated the assistance of physics can enable deep learning with a limited amount of measured silicon data. The idea of appending theoretically generated C–V with practical C–V leads to the physics-assisted ML model. The theoretical C–V is calculated in MATLAB.47 The baseline MLP and our presented advanced deep learning models are implemented using Python 3.7.7,48 TensorFlow49,50 2.0 with Keras, Numpy 1.17.0,51 Scipy 1.6.2,52 Matplotlib,53 Scikit-Learn 0.23.1,54 and Pandas.55

Equation 11 measures the closeness of fitting parameters with the original curve, and $Y_{\text{mean}}$ is the average value of the original output. Values approaching 1 mean curve tending toward their true value. The concept of the validation set is employed in such a way that the main training data set $D$ is split into two parts $D_1$ and $D_2$ i.e., $D = D_1 \cup D_2$. There is no overlapping of the validation set $D_2$ with the actual training data set $D_1$, i.e., $D_1 \cap D_2 = \emptyset$. It has been observed that early stopping using validation loss makes our model stable and robust. The loss function depends on the number of training epochs. The problem of over and underfitting is very common in regression types of prediction. In general, advanced and complex ML algorithms eliminate the underfitting problem. To mitigate the overfitting problem and obtain the best-fit prediction, the

$$\text{Equation 11}$$

$$\begin{align*}
R^2 &= 1 - \frac{\sum_{i=1}^{n} (Y_{\text{original output, i}} - Y_{\text{predicted output, i}})^2}{\sum_{i=1}^{n} (Y_{\text{original output, i}} - Y_{\text{mean}})^2} \\
\text{RMSE} &= \sqrt{\frac{\sum_{i=1}^{n} (Y_{\text{original output, i}} - Y_{\text{predicted output, i}})^2}{n}}
\end{align*}$$

Figure 5. Machine learning algorithms: (a) multilayered perceptron,56 (b) semisupervised multitasked learning,37,57 (c) and transfer learning.39,58

Figure 6. MLP architecture used as a baseline model: (a) architecture, (b) symbol definition, and (c) symbol pattern definition.
callback class of Tensorflow provides the concept of early stopping.

3.1. Performance of MLP Model. Since MLP is a conventional supervised learning ML model\(^59\) that requires a huge set of training data for a specific task, it is not going to be an attractive algorithm for designing a unified, robust model, especially when a limited amount of data is available. In Figure 6a, the baseline feedforward model is trained entirely on experimental silicon data. The two stages of the hidden layer are incorporated with the input and output layer. Figure 6b defines the nomenclature of various symbols that are used in the following set of equations. Figure 6c is the associated equations and symbol definitions, respectively. A structure of 50 neurons in each hidden layer is used for training.

### Table 3. Performance Metrics

| data   | frequency (kHz) | neural unit | RMSE (ideal value = 0) | \(R^2\)-score (ideal value = 1) |
|--------|-----------------|-------------|-------------------------|---------------------------------|
|        |                 |             | MLP | SSMTL | TL | MLP | SSMTL | TL |
| train  | 1, 100          | 50          | 0.3525 | 0.2687 | 0.7060 | 0.9946 | 0.9968 | 0.9782 |
| test   | 3, 5, 10, 50    | 50          | 3.1387 | 1.1890 | 1.3752 | 0.6108 | 0.9442 | 0.9253 |
| test   | 3                | 20          | 2.2544 | 1.6389 | 2.0799 | 0.7750 | 0.8811 | 0.8085 |
| test   | 5                |             | 3.2344 | 1.9014 | 2.3588 | 0.5414 | 0.8415 | 0.7561 |
| test   | 10               |             | 4.2492 | 2.0010 | 2.0064 | 0.2719 | 0.8385 | 0.8377 |
| test   | 50               |             | 1.8813 | 1.0769 | 1.2184 | 0.8725 | 0.9582 | 0.9465 |
| test   | 3                | 30          | 2.2358 | 1.2458 | 1.8533 | 0.7787 | 0.9313 | 0.8479 |
| test   | 5                |             | 3.2173 | 1.4405 | 2.0942 | 0.5462 | 0.9090 | 0.8077 |
| test   | 10               |             | 4.2535 | 1.6111 | 2.4261 | 0.2704 | 0.8953 | 0.7627 |
| test   | 50               |             | 2.1803 | 0.8155 | 1.7505 | 0.8287 | 0.9760 | 0.8896 |
| test   | 3                | 40          | 2.1837 | 1.5527 | 2.2032 | 0.7889 | 0.8933 | 0.7851 |
| test   | 5                |             | 3.1399 | 1.9239 | 2.1482 | 0.5678 | 0.8377 | 0.7977 |
| test   | 10               |             | 4.1207 | 1.8766 | 2.5206 | 0.3153 | 0.8580 | 0.7438 |
| test   | 50               |             | 1.9595 | 1.0005 | 1.7931 | 0.8616 | 0.9639 | 0.8841 |
| test   | 3                | 50          | 2.2633 | 1.3807 | 1.6781 | 0.7732 | 0.9156 | 0.8753 |
| test   | 5                |             | 3.2599 | 1.4223 | 1.5566 | 0.5341 | 0.9112 | 0.8938 |
| test   | 10               |             | 4.3049 | 1.2230 | 1.2315 | 0.2527 | 0.9397 | 0.9388 |
| test   | 50               |             | 2.2637 | 0.4769 | 0.8994 | 0.8153 | 0.9918 | 0.9708 |
| test   | 3                | 60          | 2.2934 | 1.5667 | 1.8708 | 0.7671 | 0.8913 | 0.8450 |
| test   | 5                |             | 3.2755 | 1.8201 | 1.7370 | 0.5297 | 0.8548 | 0.8677 |
| test   | 10               |             | 4.2911 | 1.5416 | 1.6526 | 0.2575 | 0.9042 | 0.8899 |
| test   | 50               |             | 2.2666 | 0.6480 | 1.0909 | 0.8149 | 0.9849 | 0.9571 |

Figure 7. \(C-V\) response of ML models on 1 kHz and 100 kHz, 0.0050 mm\(^2\)/RCA/Wet/E-Gun, training data for (a) baseline MLP, (b) SS-MTL, and (c) transfer learning. 50 neurons in a hidden layer is used.

Figure 8. Scatter plot of 50 neuron baseline MLP on the test data for all process parameters, i.e. frequency/area/clean type/oxide deposition/metal deposition: (a) 3, (b) 10, and (c) 50 kHz data.
and convergence of a model depend on the range of the data set. Under such circumstances, it becomes important to perform some preprocessing. MinMaxScaler form Scikit flows is used for scaling, and this method does not change the significance of outliers. The rectified linear unit (ReLU) type of activation function provides the nonlinearity in the MLP. The output signals coming from the previous hidden layer act as input for the current hidden layer. To make the model robust for any generalized data, no tuning of hyperparameters is done in accordance with the test data set. In MLP, the validation set consists of the 1 and 100 kHz of 0.01 mm² area. The remaining data of 1 and 100 kHz for different device areas are used for training purposes. The equation shown in Figure 6a gives the value of capacitance for different frequencies range. It has been found from Table 3 that RMSE and $R^2$ score for training are 0.3525 and 0.9946, respectively. Figure 7a–c shows the training set fitting for MLP, SSMTL, and TL, respectively.

The MLP model is exclusively trained on low frequency (1 kHz) and high frequency (100 kHz) of the measured C–V data, and the training response is shown in Figure 7a. The scatter plot of MLP for the different frequencies on test data is shown in Figure 8. The performance for 3, 5, 10, and 50 kHz is not satisfactory. This is due to the fact that the model is trained on 1 and 100 kHz C–V only using purely supervised learning.

No physical data are used in this algorithm, and the model is also not exposed to the C–V data of different intermediate frequencies.

Figures 7a and 8 show the predicted C–V for the training set and the scatter plot for the test set for the MLP model. It is observed that the MLP is not able to utilize the training knowledge to predict the test data sets accurately. Thus, we need the knowledge of physical equations to assist the machine learning prediction, especially when the training data is insufficient or costly to be collected. Figure 8 suggests that the midrange capacitance at 3 kHz and low capacitance values at 50 kHz do not give a satisfactory result. In the case of 10 kHz, it has been seen that the prediction for both low and midrange capacitance is worst in Figure 8. Table 3 shows the $R^2$ score and RMSE values for the MLP model. Considering the case of 50 neurons, the RMSE for 3, 5, 10, and 50 kHz is 2.2633, 3.2599, 4.3049, and 2.2637, respectively. The $R^2$ scores, which gives the idea of fitting, for 3, 5, 10, and 50 kHz are 0.7732, 0.5341, 0.2527, and 0.8153, respectively. The prediction for 10 kHz of the C–V characteristics possesses a less satisfactory response. This can be visualized from the scatterplot of Figure 8. The prediction of the 10 kHz data is more difficult because the interpolation in frequency is far from the training set frequency, i.e., 1 kHz and 100 kHz frequency data. In addition, the MLP model is not able to map the effect of carrier dynamics at intermediate frequencies due to the fact that no physical model is available for any assistance. We have also tested the supported vector regression (SVR) and random forest (RF) as a baseline model. The SVR and RF are less effective than MLP. The SVR test set RMSE and $R^2$ are 3.0788 and 0.6252, respectively. The RF is slightly better than SVR. Nevertheless, the trend in the data is not predicted well, and thus flat segments are seen in the predicted vs true value plot. RF RMSE and $R^2$ are 2.5503 and 0.7431, respectively. The hyperparameters follow the sklearn library default except epsilon = 0.2 for SVR and max_depth = 2 for RF.

3.2. Performance of the SSMTL Model

In the previous section, we have observed that the baseline MLP does not respond well to intermediate frequency data. The physics behind the C–V for intermediate frequencies is the same, i.e., the change of the frequency in the sweeping voltage signal inversion layer affects the amount of the minority electrons that can be laterally transferred to the region beneath the top electrode, and this further affects the capacitance across the semiconductor. Figure 9a shows the layered architecture of the SSMTL model. In this model, the model parameter for the first hidden layer is fed with 1 and 100 kHz of real silicon C–V data and 1, 3, 5, 10, 50, and 100 kHz of theoretical C–V data.

If a human brain is trained for a task, the knowledge of that task can automatically be utilized for similar types of new tasks. Similarly, the knowledge of all frequency theoretical C–V present in the first hidden layer is being shared with the measured data, and its pictorial representation is given by Figure 9a. Since the knowledge of theoretically generated C–V is to assist the measured C–V in training the model, it becomes necessary to take some data from both practical and theoretical data as a validation set. One of the device fabrication process parameters is the area of the electrodes, and we have six different areas. Out of these 6, we use 0.01 mm² area data of 1 and 100 kHz from the measured data set and 0.01 mm² area of all frequencies from the theoretical C–V data set as the validation set. The remaining data are used for training and testing purposes, and there is no overlapping of training, validation, and test data set. The metrics that show the performance of the SSMTL are RMSE and $R^2$, and their
values on training data are 0.2687 and 0.9968, respectively. It has been observed through Table 3 that the baseline MLP model is not efficient if compared with SSMTL. The fitting response on training data is shown in Figure 7b.

Figure 10 shows the scatter plot of SSMTL on the test data. Compared with the baseline model, it has been noticed that the overall RMSE and $R^2$ on test data are improved by 62.12% and 54.58%, respectively. It is evident from Table 3 that the $R^2$ score and RMSE for all the individual frequencies are significantly improved. Taking the case of 50 neurons, the RMSE for 3, 5, 10, and 50 kHz is 1.3807, 1.4223, 1.2230, and 0.4769, respectively, and the $R^2$ score for 3, 5, 10, and 50 kHz is 0.9156, 0.9112, 0.9397, and 0.9918, respectively. The induction of MOS capacitor physics with measured data, implemented with the concept of domain knowledge sharing architecture, improves the quality of the test data set prediction. If we consider the case of 50 neurons and 5 kHz test data, RMSE for SSMTL and MLP is 1.4223 and 3.2599, respectively. Thus, there is a clear improvement of 56.37% in RMSE value. Further improvement in the performance of SSMTL can be enhanced by generating a more accurate physical model. A more accurate physical model can be constructed by combining more numbers of process parameters or using more accurate device simulation. First, the inclusion of traps as one of the most important sources of variability can improve performance. Second, we can simulate the transport of carriers under illumination in the MOS capacitor by building a 2D structure with a given value of power rate. To carry out these simulations, TCAD can be one probable option.

In the field of ML, there is always a trade-off between overfitting and underfitting. An excessive number of training cycles corresponds to the overfitting issues and additionally leads to the consumption of unnecessary computational power. To tackle these situations, we set patience value as the input argument to Keras early stopping function. From Table 2, it can be clearly seen that the model attains the optimal response at 179 epoch cycles despite the fact that the maximum epoch is set to 8000 epochs. This approach prevents overfitting and saves a significant amount of computational power, ergo making the ML-based compacted model a state-of-the-art statistical tool. By analyzing the metrics from Table 3, it has been seen that 10 kHz of $C-V$ data has less satisfactory performance. The prediction error of 1.2230 for 50 neurons and 10 kHz data is more pronounced in the low capacitance region. Nevertheless, compared with the baseline MLP, there is still a drastic improvement in the $R^2$ score and RMSE values. Therefore, we can see that the interpolation in frequency is still better at 10 kHz using SSMTL. This indicates that the phenomenon of change in inversion charge under the effect of light is well articulated in the equation-based model described in the methodology section, even if TCAD has not been used in this work. It is worth mentioning that the accuracy of the physical model lies in the coverage of different aspects of device physics. If the energy of the incident photon is equal to or greater than the bandgap of the semiconductor, the process of optical absorption takes place, further generating electron–hole pairs. Specific to our experiment, electrons are the minority charge carrier and responsible for the transition from depletion to inversion mode.

The essence of physics is reflected in the equation-based physical approach though further improvement in the physical model, such as including the effect of the process parameters, can be valuable. Finally, we want to indicate that, in conventional supervised multitask learning, the size of the theoretical and measured data sets should be the same.
Nonetheless, semisupervised learning gives an advantage that two data sets with different sizes can be accommodated in a model.\textsuperscript{57} Specific to our experiment, the shape of theoretical data is (144288,7) while the measurement data is (4896,7). Therefore, using semisupervision makes the MTL algorithm become more versatile, and this benefits general industrial applications.

### 3.3. Performance of the Transfer Learning Model.

For a user, an ML model is just a black box. It is desired to reduce the measurement and fabrication loading in the semiconductor industry since process development is costly in foundries. This aspect gives the idea of a transfer learning-based model.\textsuperscript{61} The structure of TL is shown in Figure 11. For implementing the complex task, a deep network is required.\textsuperscript{62} The deeper the network, the more computing resources will be needed. Thus, to increase the robustness and productivity, transfer learning using the previously trained model can be illustrative. Here the pretrained model is trained on all theoretical frequency $C-V$ data. Afterward, the optimized weights and biases are transferred to the main model. The main model uses only real silicon $C-V$ data. Hyperparameters with their values are shown in Table 2.

It has been observed from the performance metric Table 3 that the $R^2$ score and RMSE at the test set for all the individual frequencies are improved relative to MLP models. In Table 3, the neuron number of the transferred layer, i.e., $H^{1,T}$ in Figure 11, is fixed at 50, and the neuron number of the retrained layer, i.e., $H^{2,P}$ in Figure 11, is varied from 20 to 60. $H^{2,T}$ is fixed at 50. Considering 50 neurons of TL model, RMSEs for 3, 5, 10, and 50 kHz are $1.6781, 1.5566, 1.2315,$ and $0.8994$, respectively, and $R^2$ scores for 3, 5, 10, and 50 kHz are $0.8753, 0.8938, 0.9388,$ and $0.9708$, respectively. Thus, it has been evident that the TL-based compact model performs more accurately at the test set. The assistance provided by the theoretical model enhances the conditional probability distribution of the main model. The domain knowledge of the predefined and main model is not exactly the same, but the task of $C-V$ prediction for different frequencies is the same.

Through the transfer of domain knowledge from predefined to the main model, the TL-based compact model outclasses the traditional baseline MLP. The performance of SS-MTL is slightly more accurate as compared to TL if the case of 50 neurons in the hidden layer is used. It has been observed from the performance metric Table 3 that the $R^2$ score and RMSE at the test set for all the individual frequencies are improved relative to MLP models. In Table 3, the neuron number of the transferred layer, i.e., $H^{1,T}$ in Figure 11, is fixed at 50, and the neuron number of the retrained layer, i.e., $H^{2,P}$ in Figure 11, is varied from 20 to 60. $H^{2,T}$ is fixed at 50. Considering 50 neurons of TL model, RMSEs for 3, 5, 10, and 50 kHz are $1.6781, 1.5566, 1.2315,$ and $0.8994$, respectively, and $R^2$ scores for 3, 5, 10, and 50 kHz are $0.8753, 0.8938, 0.9388,$ and $0.9708$, respectively. Thus, it has been evident that the TL-based compact model performs more accurately at the test set. The assistance provided by the theoretical model enhances the conditional probability distribution of the main model. The domain knowledge of the predefined and main model is not exactly the same, but the task of $C-V$ prediction for different frequencies is the same.

Through the transfer of domain knowledge from predefined to the main model, the TL-based compact model outclasses the traditional baseline MLP. The performance of SS-MTL is slightly more accurate as compared to TL if the case of 50 neurons in the hidden layers is examined. In general, the relative strengths of SSMTL and TL show case-by-case phenomenon and are roughly at a similar level and all above the baseline MLP models. The advantage of SSMTL over TL is that there is no need to determine the fixed epoch number, i.e., 30 in this work, in the training of the main model for SS-MTL. Due to the transfer of domain knowledge, the main model requires fewer epochs for training. Only 30 cycles are required to get the desired results from the main model. The RMSE value for 50 neurons presented in Table 3 shows that the 3 kHz $C-V$ prediction is less accurate compared to the $C-V$ at other frequencies. This can be observed from the scatter plot of Figure 12, where the predicted capacitance values slightly more deviate from their measured values, but the $R^2$ score of 0.8753 is still satisfactory. One of the potential further advancements in TL is the application of the diffused model with TL.\textsuperscript{63} The spatial and temporal dependency of data is captured with the diffusion of convolutional neural networks and recurrent neural network to upgrade the model predicting behavior.

On the other hand, performance-wise, meta transfer learning is also an advanced concept which further reduces the requirement of large labeled data and increasing the overall performance.\textsuperscript{64} The training performance of baseline MLP on low frequency and high-frequency test data is shown in Figure 7a. An RMSE of 0.3525 and $R^2$ of 0.9946 are obtained. MLP is
a commonly used algorithm in a regression-type problem. Nevertheless, when it comes to the unseen or temporal type of data, it can become ineffective. This behavior can undoubtedly be observed in Table 3. The RMSE and $R^2$ values are severely affected in the test data of intermediate frequencies. The performance measures of transfer learning are shown in Table 3. The architecture of the pretrained model consists of two hidden layers and is trained with all frequencies of theoretical data. Afterward, the weight and bias are transferred to a new model that predicts the electrical characteristics of intermediate frequencies. The batch size and epoch cycle of the pretrained model are tuned in such a way that the processor power and fitting characteristics are not compromised. From Table 3 and Figure 12, it can be seen that, after implementing transfer learning and SSMTL, there is a wide improvement in the performance metrics. The baseline MLP model has good performance on training data, but the model applicability lies on the test data performance. By including the concept of theoretical $C$–$V$, the physics-assisted SSMTL and TL models outclass the baseline model.

3.4. Application of ML Models in Device Fabrication. To make the model robust with broad applications, it is illustrative to conjugate the concept of intelligent manufacturing with SPICE modeling. Our model can easily be used in understanding the effect of various fabrication processes on device performance. In our experiment, we have considered the effect of different cleaning techniques, different thermal deposition techniques, and different metal electrode deposition techniques. Figure 13 shows the model $C$–$V$ characteristics variation for one type of process, i.e., the types of thermal deposition. From Figure 13, taking RCA cleaning, 0.01 mm$^2$, E-gun metal deposition, and 50 kHz data, it is clear that the effect of process variation on $C$–$V$ characteristics can easily be studied by our prescribed architecture of the ML models. Finally, the usability of the method here can lie in that measuring less CV or IV curves can expedite the process development time and cycles, especially in the early development stage of semiconductor devices. It is also illustrative to discuss the potential of using the same approach here to predict capacitance in different scenarios. Using the approach here, predicting the missing capacitance values at some voltages for a device will be quite accurate since this is mostly an interpolation problem. In case we have to predict the capacitance values across devices with different process conditions, the accuracy will be degraded, and the physics assistance has no effect since the physical model does not encompass process information such as wafer cleaning.

4. CONCLUSION

The prediction of device performance as the result of complex semiconductor manufacturing procedures is always desired. In this work, we demonstrate two advanced data-driven physics-assisted ML-based models that are effectively used to map the device $C$–$V$ to oxidation and wafer cleaning procedures. We observe that semisupervised multitask learning (SSMTL) and transfer learning (TL) achieve better performance compared to the standard MLP models. The improved data-driven models are based on the sharing or transferring of weights and biases of some layers in the ML models. The hybridization utilizes the theoretical $C$–$V$ of 1, 3, 5, 10, 50, and 100 kHz, and the measured $C$–$V$ of 1 and 100 kHz are used to predict the measured $C$–$V$ for intermediate frequencies at 3, 5, 10, and 50 kHz. The samples are fabricated by considering different fabrication conditions targeted at constructing data-driven compact device models with process-aware capability.

The most significant advantage of our presented data-driven model is that it can smoothly expand to any number of process parameters. In addition, the knowledge sharing and transferring in SSMTL and TL setups leads to relaxed requirements on the amount of data collected in semiconductor manufacturing, which can be costly. This makes our presented model more practical in production lines. Semisupervised multitask learning (SSMTL) and transfer learning (TL) algorithms are tested on various unseen $C$–$V$ data, and $R^2$ values of 0.9442 and 0.9253 are obtained for the test set data for SSMTL and TL, respectively. We believe that this work will be important for future smart semiconductor manufacturing, in which the massive chemical reactions in semiconductor processes have to be accurately modeled to predict the final device characteristics.

# AUTHOR INFORMATION

Corresponding Author
Albert S. Lin — Institute of Electronics Engineering, National Yang Ming Chiao-Tung University, Hsinchu City 30010 Taiwan, R.O.C.; orcid.org/0000-0001-9707-7466; Email: ldtd5746@gmail.com

Authors
Sparsh Pratik — Institute of Electronics Engineering, National Yang Ming Chiao-Tung University, Hsinchu City 30010 Taiwan, R.O.C.; orcid.org/0000-0002-2927-7121
Po-Ning Liu — Institute of Electronics Engineering, National Yang Ming Chiao-Tung University, Hsinchu City 30010 Taiwan, R.O.C.; orcid.org/0000-0003-3135-7099
Jun Ota — Institute of Electronics Engineering, National Yang Ming Chiao-Tung University, Hsinchu City 30010 Taiwan, R.O.C.; Imari Crystal Engineering Section, Crystal Engineering Department, Technology Division, SUMCO Corporation, Saga 849-0506, Japan; orcid.org/0000-0001-9860-1186
Yen-Liang Tu — Institute of Electronics Engineering, National Yang Ming Chiao-Tung University, Hsinchu City 30010 Taiwan, R.O.C.; Present Address: Y.-L.T.: Applied Material, Taiwan; orcid.org/0000-0003-0613-343X
Guang-Wen Lai — Institute of Electronics Engineering, National Yang Ming Chiao-Tung University, Hsinchu City 30010 Taiwan, R.O.C.; Present Address: G.-W.L.: Taiwan Semiconductor Manufacturing Corporation (TSMC), Taiwan; orcid.org/0000-0002-2929-0430
Ya-Wen Ho — Institute of Electronics Engineering, National Yang Ming Chiao-Tung University, Hsinchu City 30010 Taiwan, R.O.C.; orcid.org/0000-0001-9707-7466
Zheng-Kai Yang — Institute of Electronics Engineering, National Yang Ming Chiao-Tung University, Hsinchu City 30010 Taiwan, R.O.C.; orcid.org/0000-0001-7116-3403
Tejender Singh Rawat — Institute of Electronics Engineering, National Yang Ming Chiao-Tung University, Hsinchu City 30010 Taiwan, R.O.C.; orcid.org/0000-0001-7894-0463

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.1c05552

Notes
The authors declare no competing financial interest.
ACKNOWLEDGMENTS

This work was supported by the Ministry of Science and Technology, Taiwan, under grant number MOST 109-2221-E-009-060. The authors acknowledge the support of the Taiwan Semiconductor Research Institute (TSRI), Taiwan.

REFERENCES

(1) Meijer, P. B. L. Neural Network Applications in Device and Subcircuit Modelling for Circuit Simulation; Ph.D. Thesis, Technische Universiteit Eindhoven, 1996.

(2) Creech, G. L.; Paul, B. J.; Lesnäät, C. D.; Jenkins, T. J.; Calcaterra, M. C. Artificial neural networks for fast and accurate EM-CAD of microwave circuits. IEEE Trans. Microwave Theory Technol. 1997, 45 (5), 794–802.

(3) Klemme, F.; Prinz, J.; Santen, V. M.; Henkel, J.; Amrouch, H. Modeling Emerging Technologies using Machine Learning: Challenges and Opportunities. IEEE/ACM International Conference on Computer Aided Design (ICCAD), Nov. 2–5, 2020; pp 1–9.

(4) Hammouda, H.; Mhiri, M.; Gafsi, Z. d.; Besbes, K. Neural-Based Models of Semiconductor Devices for SPICE Simulator. Am. J. Appl. Sci. 2008, 5 (4), 385.

5) Zhang, Z.; Wang, R.; Chen, C.; Huang, Q.; Wang, Y.; Hu, C.; Wu, D.; Wang, J.; Huang, R. New-Generation Design-Technology Co-Optimization (DTCO): Machine-Learning Assisted Modeling Framework. arXiv.org, 2019, 1904.10269.

(6) Lu, D. D.; Dunga, M. V.; Niknejad, A. M.; Hu, C.; Liang, F.-X.; Hung, W.-C.; Lee, J.-W.; Hsu, C.-H.; Chiang, M.-H. Compact Device Models for FinFET and Beyond. arXiv.org, 2020, 2005.02580.

(7) Liang, W.; Yang, X.; Loiseau, A.; Mitra, S.; Gauthier, R. Novel ESF Compact Modeling Methodology using Machine Learning Techniques. 2020 42nd Annual EOS/ESD Symposium (EOS/ESD); Sept. 13–18, 2020; pp 1–7.

(8) Nicollian, E. H.; Brews, J. R. MOS (metal oxide semiconductor) physics and technology; Wiley: New York, 1982.

(9) Taur, Y.; Ning, T. H. Fundamentals of Modern VLSI Devices; Cambridge University Press, 2009.

(10) Krishnamurthy, S.; Lin, X.-W.; Mahotin, Y.; Mukherjee, P.; Roger, F.; Ryles, R.; Uppal, S.; Cuthbertson, A. Si-Based Process Aware SPICE Models for Statistical Circuit Analysis. NSTI Nanotech. 2008, 897–900.

(11) Braun, G.; Dasarapu, V. K.; Karmarkar, A. P.; Krishnamurthy, S.; Lin, X.-W.; Saha, A. R. Process Aware Compact Model Parameter Extraction for 45 nm Process. NSTI Nanotech. 2008, 833–836.

(12) Zhong, R. Y.; Xu, X.; Klotz, E.; Newman, S. T. Intelligent Manufacturing in the Context of Industry 4.0: A Review. Engineering 2017, 3 (5), 616–630.

(13) Chen, B.; Wan, J.; Shu, L.; Li, P.; Mukherjee, M.; Yin, B. Smart Factory of Industry 4.0: Key Technologies, Application Case, and Challenges. IEEE Access 2018, 6, 6505–6519.

(14) Espadinha-Cruz, P.; Godina, R.; Rodrigues, E. M. G. A Review of Data Mining Applications in Semiconductor Manufacturing. Processes 2021, 9 (2), 305.

(15) Chauhan, Y. S.; Venugopalan, S.; Karim, M. A.; Khandelwal, S.; Paydavosi, N.; Thakur, P.; Niknejad, A. M.; Hu, C. C. In BSIM — Industry standard compact MOSFET models. 2012 Proceedings of the ESSCIRC (ESSCIRC) 2012, 30–33 17–21 Sept. 2012.

(16) Popel, M.; Tomkova, M.; Tomek, J.; Kaiser, L.; Uszkoreit, J.; Bojar, O.; Zabokrtsky, Z. Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals. Nat. Commun. 2020, 11 (1), 4381.

(17) Isiyan, F. O.; Folajimi, Y. O.; Ojokoh, B. A. Recommendation systems: Principles, methods and evaluation. Egyptian Informatics Journal 2015, 16 (3), 261–273.

(18) Kim, Y.; Jeon, J.; Jang, Y.; Park, Y.; Yang, G.; Park, Y.; Yoo, M.; Chung, C. In Compact process and layout aware model for variability optimization of circuit in nanoscale CMOS. 2010 International Conference on Simulation of Semiconductor Processes and Devices 2010, 141–144 6–8 Sept. 2010.

(19) Sponton, L. From Manufacturing Variability to Process-aware Circuit Simulation; ETH, 2007.

(20) Overstall, A. M.; Woods, D. C.; Martin, K. J. Bayesian prediction for physical models with application to the optimization of the synthesis of pharmaceutical products using chemical kinetics. Computational Statistics & Data Analysis 2019, 132, 126–142.

(21) Read, J. S.; Jia, X.; Willard, J.; Appling, A. P.; Zwart, J. A.; Oliver, S. K.; Karpatne, A.; Hansen, G. J. A.; Hanson, P. C.; Watkins, W.; Steinbach, M.; Kumar, V. Process-Guided Deep Learning Predictions of Lake Water Temperature. Water Resour. Res. 2019, 55 (11), 9173–9190.

(22) Huang, C. Y.; Fu, S. M.; Parashar, P.; Chen, C. H.; Akbar, C.; Lin, A. S. Intelligent Manufacturing: TCAD-Assisted Adaptive Weighting Neural Networks. IEEE Access 2018, 6, 78402–78413.

(23) Chen, C. H.; Parashar, P.; Akbar, C.; Fu, S. M.; Syu, M.; Lin, A. Physics-Prior Bayesian Neural Networks in Semiconductor Processing. IEEE Access 2019, 7, 130168–130179.

(24) Kim, Y.; Myung, S.; Ryu, J.; Jeong, C.; Kim, D. S. In Physics-augmented Neural Compact Model for Emerging Device Technologies. 2020 International Conference on Simulation of Semiconductor Processes and Devices (SIPSAD) 2020, 257–260.

(25) Gao, X.; Huang, A.; Trask, N.; Reza, S. In Physics-Informed Graph Neural Network for Circuit Compact Model Development. 2020 International Conference on Simulation of Semiconductor Processes and Devices (SIPSAD) 2020, 359–362.

(26) Li, M.; Irsoy, O.; Cardie, C.; Xing, H. G. Physics-Inspired Neural Networks for Efficient Device Compact Modeling. IEEE Journal on Exploratory Solid-State Computational Devices and Circuits 2016, 2, 44–49.

(27) Lin, A. S.; Pratik, S.; Ota, J.; Rawat, T. S.; Huang, T. H.; Hsu, C. L.; Su, W. M.; Tseng, T. Y. A Process-Aware Memory Compact-Device Model Using Long-Short Term Memory. IEEE Access 2021, 9, 3126–3139.

(28) Han, D.; Liu, Q.; Fan, W. A new image classification method using CNN transfer learning and web data augmentation. Expert Systems with Applications 2018, 95, 43–56.

(29) Imoto, K.; Nakai, T.; Ike, T.; Haruki, K.; Sato, Y. A CNN-Based Transfer Learning Method for Defect Classification in Semiconductor Manufacturing. IEEE Transactions on Semiconductor Manufacturing 2019, 32 (4), 455–459.

(30) Liu, Z.; Jiang, M.; Luo, T. Leverage electron properties to predict phonon properties via transfer learning for semiconductors. Sci. Adv. 2020, 6 (45), No. eaab1356.

(31) Jha, D.; Choudhary, K.; Tavazza, F.; Liao, W.-k.; Choudhary, A.; Campbell, C.; Agrawal, A. Enhancing materials property prediction by leveraging computational and experimental data using deep transfer learning. Nat. Commun. 2019, 10 (1), 5316.

(32) Li, X.; Dan, Y.; Dong, R.; Cao, Z.; Niu, C.; Song, Y.; Li, S.; Hu, J. Computational Screening of New Perovskite Materials Using Transfer Learning and Deep Learning. Appl. Sci. 2019, 9 (24), 5510.

(33) Yamada, H.; Liu, C.; Wu, S.; Koyama, Y.; Ju, S.; Shiomji, J.; Morikawa, J.; Yohida, R. Predicting Materials Properties with Little Data Using Shotgun Transfer Learning. ACS Cent. Sci. 2019, 5 (10), 1717–1730.

(34) Park, C.; Kim, Y.; Park, Y.; Kim, S. B. Multitask learning for virtual metrology in semiconductor manufacturing systems. Computers & Industrial Engineering 2018, 123, 209–219.

(35) Xiong, J.; Zhu, Y.; He, J. Machine Learning for VLSI Chip Testing and Semiconductor Manufacturing Process Monitoring and Improvement. In Machine Learning in VLSI Computer-Aided Design; Elfadil, I. M.; Boning, D. S.; Li, X., Eds.; Springer International Publishing: Cham, 2019; pp 233–263.

(36) Zhang, Y.; Yeung, D.-Y. Semi-Supervised Multi-Task Regression. Machine Learning and Knowledge Discovery in Databases; Buntine, W.; Grobelnik, M., Madlenic, D.; Shawe-Taylor, J., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2009; pp 617–631.

(37) Ruder, S. An Overview of Multi-Task Learning in Deep Neural Networks. arXiv.org, 2017, 1706.05098.
(38) Jun, O.; Sparsh, P.; Po, N. L.; Lai, G.-L.; Tu, Y.-L.; Lin, A. Physics-Assisted Compact Device Modeling using Semi-supervised Multi-task and Transfer Learning. In International Conference on Manufacturing Technologies (ICMT 2021); IOP Conference Series: Materials Science and Engineering; Georgia Institute of Technology, Atlanta, GA, USA, 2021.

(39) Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A Comprehensive Survey on Transfer Learning. Proc. IEEE 2021, 109 (1), 43–76.

(40) Yang, Q. In An Introduction to Transfer Learning, Advanced Data Mining and Applications; Tang, C., Ling, C. X., Zhou, X., Cercone, N. J., Li, X., Eds.; Springer: Berlin, Heidelberg, 2008; pp 1–1.

(41) Tu, Y. L. Intelligent Manufacturing of Dry Oxide MOSCAP Device by Using Ensemble Learning; M.S. Thesis, National Chiao Tung University, Hsinchu, Taiwan, 2020.

(42) Lai, G.-W. Process Aware Machine Learning Model of Mos Capacitor by Using Multi-Task Learning; M.S. Thesis, National Chiao Tung University, Hsinchu, Taiwan, 2020.

(43) Poutinen, W. K. A. D. A. Cleaning solution based on hydrogen peroxide for use in silicon semiconductor technology. RCA Rev. 1970, 31, 187–205.

(44) Han, L.; Yao, S.; Theuwissen, A. J. P. A Charge Transfer Model for CMOS Image Sensors. IEEE Trans. Electron Devices 2016, 63 (1), 32–41.

(45) Kuhn, M.; Johnson, K. Applied Predictive Modeling 2013, DOI: 10.1007/978-1-4614-6849-3.

(46) Dipanjan Sarkar, R. B., Tamogna, G. Hands-On Transfer Learning with Python: Implement advanced deep learning and neural network models using TensorFlow and Keras; Packt Publishing, 2018.

(47) Maths Works. Matlab. 11.16.2.1 (R2019b)); The MathWork-score.

(48) Rossum, V.; Guido, Drake, F. L. Python 3 Reference Manual; CreateSpace: Scotts Valley, CA, 2009.

(49) Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; Kudlur, M.; Levenberg, J.; Monga, R.; Moore, S.; Murray, D. G.; Steiner, B.; Tucker, P.; Vasudevan, V.; Warden, P.; Wicke, M.; Yu, Y.; Zheng, X. TensorFlow: A system for large-scale machine learning. arXiv.org, 2016, 1605.08695.

(50) Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G. S.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Goodfellow, I.; Harp, A.; Irving, G.; Isard, M.; Jozefowicz, R.; Jia, Y.; Kaiser, L.; Kudlur, M.; Levenberg, J.; Mané, D.; Schuster, M.; Monga, R.; Moore, S.; Murray, D.; Olah, C.; Schlsens, J.; Steiner, B.; Sutskever, I.; Talwar, K.; Tucker, P.; Vanhoucke, V.; Vasudevan, V.; Viégas, F.; Vinyals, O.; Warden, P.; Wattenberg, M.; Wicke, M.; Yu, Y.; Zheng, X. TensorFlow: Large-scale machine learning on heterogeneous systems. arXiv.org, 2015, 1603.04467.

(51) Harris, C. R.; Millman, K. J.; Walt, S. J. v. d.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; Kern, R.; Picus, M.; Hoyer, S.; Kerkwijk, M. H. v.; Brett, M.; Haldane, A.; Río, J. F. d.; Wiebe, M.; Peterson, P.; Gérard-Marchant, P.; Sheppard, K.; Reddy, T.; Weckesser, W.; Abbasi, H.; Gohlke, C.; Oliphant, T. E. Array programming with NumPy. Nature 2020, 585, 357–362.

(52) Jones, E.; Oliphant, T.; Peterson, P. SciPy: Open Source Scientific Tools for Python; 2001.

(53) Hunter, J. D. Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering 2007, 9 (3), 90–95.

(54) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine Learning in Python. J. Machine Learn. Res. 2011, 12 (85), 2825–2830.

(55) McKinney, W. In Data structures for statistical computing in python. Proceedings of the 9th Python in Science Conference, 2010.

(56) Baum, E. B. On the capabilities of multilayer perceptrons. Journal of Complexity 1988, 4 (3), 193–215.