ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Meta-Learned and TCAD-Assisted Sampling in Semiconductor Laser Annealing

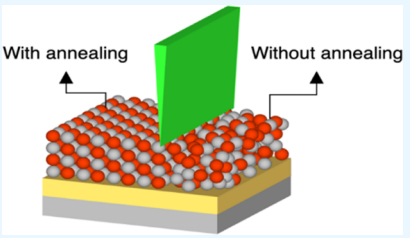
Tejender Singh Rawat, Chung Yuan Chang, Yen-Wei Feng, ShihWei Chen, Chang-Hong Shen, Jia-Min Shieh, and Albert Shihchun Lin*

Cite This: *ACS Omega* 2023, 8, 737−746

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: While applying machine learning (ML) to semiconductor manufacturing is prevalent, an efficient way to sample the search space has not been explored much in key processes such as lithography, annealing, deposition, and etching. The aim is to use the fewest experimental trials to construct an accurate predictive model. Here, we proposed a technology computer added design (TCAD)-assisted meta-learned sampling approach. The meta-learner adjusts the way of sampling in terms of how to hybridize the TCAD with ML when selecting the next sampling point. While an advanced semiconductor process is expensive, efficient sampling is indispensable. Using laser annealing as an example, we demonstrate the effectiveness of the proposed algorithm where the mean square error (MSE) at the first 100 sampling steps using TCAD-assisted meta-learned sampling is significantly lower than the pure ML approach. Besides, with reference to the pure TCAD approach, the TCAD-assisted sampling prevents the MSE degradation at 200−400 sampling steps. The proposed approach can be used in other manufacturing or even any applied machine intelligence fields.



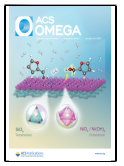
1. INTRODUCTION

The application of machine learning (ML) in semiconductor processes promotes process development and yields. In many key process steps such as lithography, deposition, annealing, and etching, the complex phenomena make purely analytical modeling difficult. In such scenarios, ML becomes a handy tool to circumvent the deficiency in the lack of proper domain-knowledge-based models. As far as the semiconductor industry is concerned, the cost of experimental trials is high. Thus, efficient sampling becomes an important task. In the case of optimization, the experiment can be conducted using a selected optimization algorithm, which has been a mature field. On the other hand, in the case of constructing an accurate model for advanced process development or predictive maintenance, data-efficient sampling in semiconductor manufacturing has been less studied compared to optimization, while sampling has been an intensively studied subject in other fields.[1−6] Prior works in data-efficient sampling in the semiconductor process are mainly in yield improvement, quality control, and predictive maintenance in a production-line setting.[7−14] On the other hand, the sampling strategies in developing key process steps in advanced technology nodes have not been studied much, and we only found limited literature.[14,15]

Laser annealing is promising in advanced semiconductor technology nodes due to its low thermal budget and capability of localized annealing. Traditional furnace annealing cannot fully eliminate defects and can result in the degradation of material properties, the precipitation of dopant substances, and the lateral dopant diffusion in wafers owing to high temperature and anneal time. In contrast, laser annealing has several advantages. First, the heating time is short. Second, the heating can be limited to a local surface region and does not affect the surrounding components. Third, it can form an abrupt profile to reduce the subthreshold leakage current. Lastly, the shape of a laser beam can be controlled meticulously in both spatial and temporal aspects.[16−18] There can be difficulty in achieving optimized sampling regarding the laser annealing condition purely by theory. The deviation from the theoretical model is mainly due to the difficulty in retrieving accurate material parameters for diffusion, melting, dopant activation, and heat transfer. Additionally, inaccurate boundary conditions, the anomalies in a polycrystalline material, and the lack of laser repetition rate and substrate heating effect models all contribute to the deviation.

Theory-assisted ML has been drawing great attention over the past few years.[19−26] The main idea is to count on prior knowledge and theory to reduce the loading of ML model training in terms of the required sample number and prediction accuracy. Various techniques have been proposed to realize

Received: September 16, 2022
Accepted: December 13, 2022
Published: December 22, 2022



https://doi.org/10.1021/acsomega.2c06000
ACS Omega 2023, 8, 737−746

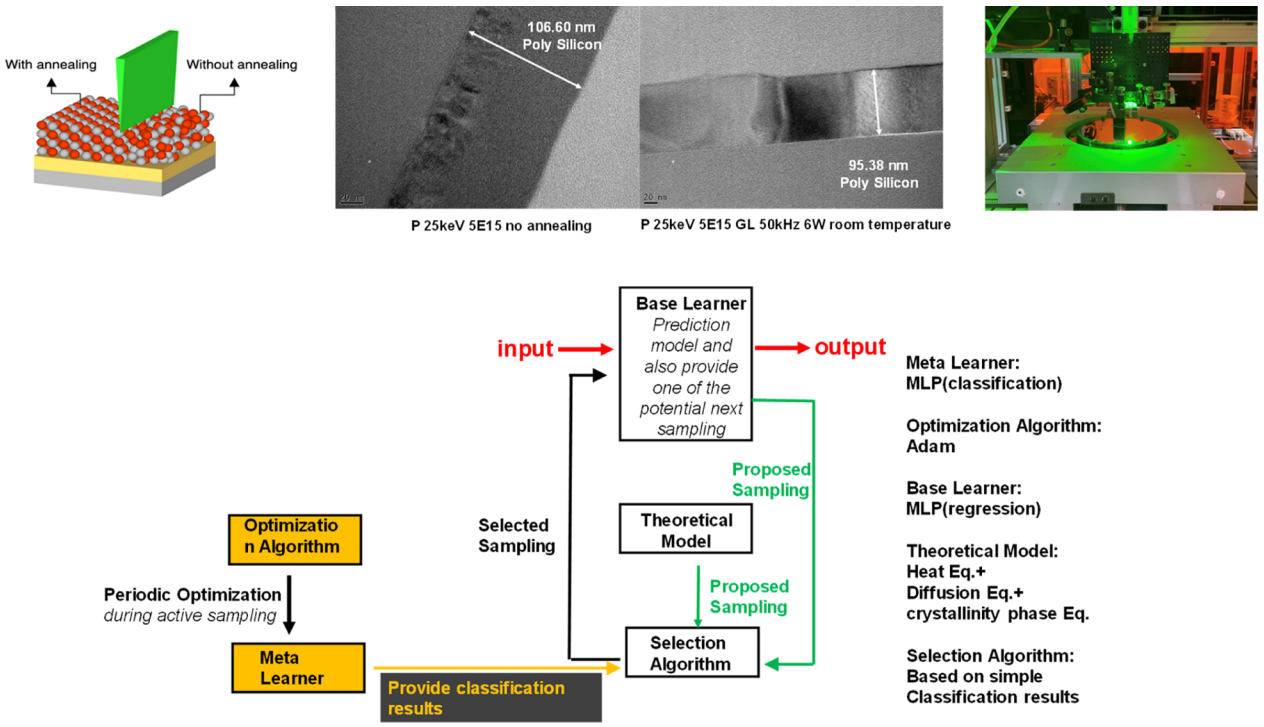


Figure 1. Illustration of meta-learned TCAD-assisted sampling algorithm. The goal is to have an accurate prediction model using the fewest sampling steps.

domain knowledge assistance, including semisupervised learning, Bayesian statistics, supervised learning incorporating domain knowledge in some context, and theory-informed reinforcement learning (RL).[23] In most of the studies in semiconductor manufacturing, the goal is to use theory-assisted prediction to reduce the required data amount, while theory-assisted sampling has not been quite studied. It should be noted that the current proposal is different from theory-assisted model-free RL since no predictive model regarding the environment is constructed in model-free RL. The proposal here is also different from theory-assisted model-based RL since RL optimizes wrt the action reward instead of the prediction accuracy on the test data set/environment.

Meta-learning[27−30] is a newly emerged field in ML where a meta-learner optimizes the base learner, enabling the paradigm of learning how to learn. It can be regarded as a nested ML model where one model learns the task while the other learns how to optimize the training of the first model. In such a scenario, many base learners will be trained and formed, and afterward, the meta-learner will learn to optimize the training process. The meta-learner can be an evolutionary algorithm or a supervised learning algorithm. In case supervised learning is used, optimization should be done wrt the input parameters of the meta-learner to locate the optimum learning strategies in the base learners. Such a nested ML model is similar to human thinking, where we do learn how to learn to improve learning efficiency in our childhood or even at later ages.

Sampling methods in ML can be categorized into passive sampling and active sampling. Passive sampling chooses samples based on the geometry of the features, while active sampling selects the next samples based on the output of the model trained by the currently available data.[31] Three methods are used in active learning: uncertainty sampling, minimizing the hypothesis space, and variance reduction.[5,32] In uncertainty sampling, several methods exist to conduct uncertainty samplings, such as least confidence and margin confidence. The least confidence method chooses data whose confidence is less than a specific number defined by the user. The margin confidence method calculates the difference between the top two most confident classes. If the difference is not large enough, the data is selected to be the next sampling point.[6,33,34] As for minimizing the hypothesis space, the aim is to locate the cases to reduce the number of possible hypotheses/models, that is, version space, induced from the given samples. Thus, an active learning algorithm manages to gain examples that can immediately minimize the number of hypotheses,[35] and the two most widely used methods are query by disagreement and query by committee.[36,37] In the variance reduction method, the active data selection is used to minimize the predictive variance of the active learning algorithm depending on Fisher information. Nevertheless, it will face a detriment that the computational complexity makes it impractical since a large number of parameters are present.[5,38−40] Hence, it has received little interest in actual applications of efficient sampling in ML problems.

2. METHOD

2.1. Algorithm. Meta-learned, technology computer added design (TCAD)-assisted sampling is illustrated in Figure 1. Essentially, the algorithm is started from sampling one sample using a semiconductor experiment. This is the standard procedure in the semiconductor industry, while in the purest computer science data mining, the training sets are collected as a whole. The cost of collecting data in real semiconductor experiments is much higher. Thus, a proper sampling strategy is required, and it is uncommon to conduct a bunch of experiments at once without further discretion. The first row in the data set is used as the starting point, and based on this first sample, an inaccurate neural network is built, serving as the

https://doi.org/10.1021/acsomega.2c06000
ACS Omega 2023, 8, 737−746

base learner in the context of meta-learning. The next sample needs to be determined based on the goal of constructing a predictively accurate model instead of locating the optimal process parameters. The way to achieve this is to know where the model predicts the worst by the currently trained ML model. Essentially, we have another useful tool in hand, which is the theoretical data for laser annealing.

In Figure 1, there are two learners: the base learner and the meta-learner. The base learner serves the purpose of prediction, which is the main purpose of essentially all of the ML tasks. In addition to being the main prediction model, the base learner also serves the purpose of suggesting the potential next sampling point during active learning. The suggestion is based on the currently constructed prediction network for the base learner, and in fact, this is the common practice in active learning or active sampling. The other suggested next sampling point will come from the theoretical model, which will be described in detail in the next section. The two suggested next sampling points need to be selected based on an algorithm. Therefore, the meta-learner serves the purpose of providing the classification results, and the selection is based on the classification results. The ML proposed next sampling point will be evaluated first. At this location, if the ML model, that is, the base learner, is more accurate according to the meta-learner, then the suggestion from the ML model will be used. Otherwise, the suggested next sampling point by the theoretical model will be used. Since the base learner is trained and updated after a certain amount of newly sampled data points are added, the meta-learner is also trained periodically during the active learning procedure.

Mathematically, we can express the ML, TCAD, and true mapping between the input process parameters x and the output metrology value y as

$$y_{\mathrm{ML}} = f_{\mathrm{ML}}(\vec{x}) \tag{1a}$$

$$y_{\mathrm{TCAD}} = f_{\mathrm{TCAD}}(\vec{x}) \tag{1b}$$

$$y_{\mathrm{true}} = f_{\mathrm{true}}(\vec{x}) \tag{1c}$$

where x is the input process parameters in a laser thermal annealing process, while y is the sheet resistance (R_{sheet}) measured. The details of the laser thermal annealing experiment can be referred to in ref.[41] In the preprocessing of the experimental and TCAD data, the sheet resistance values greater than 1000 Ω will be set to 1000 Ω since the overly large values indicate failed annealing and are not of interest. Normalization of the input and output parameters x and y to (0, 1) is conducted. There are two next sampling points proposed by ML and the TCAD model, namely, $x_{\mathrm{next,\ TCAD}}$ and $x_{\mathrm{next,ML}}$, based on whether the predicted values are far from the model mean.

$$x_{\mathrm{next,ML}} = \underset{\text{exclude 1}}{\mathrm{argmax}} |f_{\mathrm{ML}}(x) - \overline{f_{\mathrm{ML}}(x)}| \tag{2a}$$

$$x_{\mathrm{next,TCAD}} = \underset{\text{exclude 1}}{\mathrm{argmax}} |f_{\mathrm{TCAD}}(x) - \overline{f_{\mathrm{TCAD}}(x)}| \tag{2b}$$

In determining the next sampling point, we set the criterion of predicted y_{ML} <0.8 and predicted y_{TCAD} <0.8 to prevent consistent sampling on 1000 Ω data. To decide whether to use $x_{\mathrm{next,\ TCAD}}$ or $x_{\mathrm{next,ML}}$, an auxiliary classification ML model is built, and the meta-learning will adjust the model parameters of this auxiliary sampling model at each sampling step to promote better utilization of ML and TCAD models.

$$\Pi' = F_{\mathrm{classify}}(f_{\mathrm{TCAD}}(\vec{x}), f_{\mathrm{base}}(\vec{x}), \vec{x}, f_{\mathrm{true}}(\vec{x})) \tag{3}$$

$$\Pi' = \begin{cases} 0 & |f_{\mathrm{TCAD}}(\vec{x}) - y_{\mathrm{true}}| > |f_{\mathrm{ML}}(\vec{x}) - y_{\mathrm{true}}| \\ 1 & |f_{\mathrm{TCAD}}(\vec{x}) - y_{\mathrm{true}}| < |f_{\mathrm{ML}}(\vec{x}) - y_{\mathrm{true}}| \end{cases} \tag{4}$$

$$\Pi = G_{\mathrm{classify\ Fit}}(\vec{x}) \tag{5}$$

$$\vec{x}_{\mathrm{next}} = \begin{cases} \vec{x}_{\mathrm{next,TCAD}} & G_{\mathrm{classify}}(\vec{x}_{\mathrm{next,ML}}) < 0.5 \\ \vec{x}_{\mathrm{next,ML}} & G_{\mathrm{classify}}(\vec{x}_{\mathrm{next,ML}}) > 0.5 \end{cases} \tag{6}$$

Π' is the true classification between TCAD and ML accuracy derived from the sampling procedures, while Π is the fitted classification by networks. The sampling is repeated and continued until a prescribed step is achieved. The remaining data points in the data set are used to evaluate the effectiveness of our proposed meta-learning TCAD-assisted sampling method. The baselines can be a pure ML model and a pure TCAD model. It is worth mentioning that the input to the auxiliary classification ML model to decide whether to use $x_{\mathrm{next,\ TCAD}}$ or $x_{\mathrm{next,ML}}$ requires some consideration. The easiest way is to compare the accuracy of the TCAD and ML models at the previous sampling point. A further complication is to use all of the previous sampling points as the input and compare the accuracy of the ML and the TCAD model, which is the practice in this work. Alternatively, sample weights/biases and their change from the previous sampling stages of the base leaner can be used as the input. In principle, the TCAD model outperforms the ML model at an early stage of sampling and model construction, while the strength is reversed at the later stage of model construction when more data points are trained. Using sample weights/biases and their change as the input can reflect the evolved relative strength of the ML and TCAD models. Tensorflow 2.3.0,[42] Python 3.8.10 with Numpy 1.20.3 and Scipy 1.6.2,[43−46] Sentaurus Process 2016.03,[47] and Matlab R2017b[48] are used in this work.

2.2. TCAD. To compare with the results from the experiments, we use TCAD to simulate laser thermal annealing processes in the Sentaurus process. The simulations are conducted with the heat transfer equation, the dopant diffusion equation, the phase field equation, and the dopant activation model. We set the dopant activation model to solid, in which a simple solid solubility is considered.[47]

$$C_{\mathrm{A}}^{+} = \frac{C_{\mathrm{A}}^{\mathrm{SS}} C_{\mathrm{A}}}{(C_{\mathrm{A}}^{\mathrm{SS}} + C_{\mathrm{A}})} \tag{7}$$

$$C_{\mathrm{A}}^{\mathrm{SS}} = f \cdot C_{\mathrm{A}}^{\mathrm{SS0}} \tag{8}$$

where f is the multiplication factor, $C_{\mathrm{A}}^{\mathrm{ss0}}$ is the solid solubility of the dopant, and C_{A} is the total dopant concentration. The transfer matrix method (TMM), which considers the reflectance, transmittance, and absorbance for both perpendicular and parallel polarizations, is used to calculate the heat generation (G) in the heat transfer equation.[47]

$$\rho c_{\mathrm{p}} \frac{\partial T}{\partial t} = \nabla \cdot (\kappa \nabla T) + G + 30 \rho L \varphi^2 (1 - \varphi)^2 \frac{\partial \varphi}{\partial t} \tag{9}$$

$$G = I_0 \cdot g \tag{10}$$

https://doi.org/10.1021/acsomega.2c06000
ACS Omega 2023, 8, 737−746

Table 1. Comparison of Prediction Accuracy in Terms of MSE for Different Algorithms at Different Stages[a]

			pure neural network			TCAD-assisted ML			pure TCAD		
	retrain (steps)	epochs	MSE	MES_{100}	MSE_{mid}	MSE	MES_{100}	MSE_{mid}	MSE	MES_{100}	MSE_{mid}
a	5	20	0.0895	0.2642	0.0660	0.0737	0.1720	0.0971	0.0704	0.1313	0.1053
b	10	20	0.0692	0.3084	0.0700	0.0662	0.1778	0.0814	0.0722	0.1742	0.0890
c	15	20	0.0808	0.4054	0.0675	0.0918	0.3056	0.0946	0.0891	0.2276	0.0821
d	30	20	0.0790	0.4063	0.0692	0.0951	0.2901	0.1027	0.1240	0.3618	0.1106
e	5	40	0.1111	0.2398	0.1385	0.0769	0.1343	0.0791	0.0469	0.1075	0.0702
f	10	40	0.0733	0.2381	0.0616	0.0603	0.1831	0.0820	0.0768	0.1222	0.1365
g	15	40	0.0830	0.3431	0.0549	0.0867	0.2515	0.1004	0.0837	0.1461	0.1359
h	30	40	0.0940	0.4696	0.0537	0.0856	0.4822	0.0625	0.0871	0.2396	0.0840
i	5	60	0.0719	0.1753	0.0714	0.0683	0.1307	0.0993	0.0601	0.1054	0.0994
j	10	60	0.1859	0.3837	0.1013	0.0636	0.1489	0.0701	0.0855	0.1206	0.1528
k	15	60	0.0572	0.2514	0.0506	0.0644	0.1872	0.0995	0.0629	0.1336	0.0985
l	30	60	0.0685	0.2523	0.0624	0.0525	0.2198	0.0641	0.0777	0.1916	0.1082
m	5	80	0.2417	0.2466	0.5540	0.0480	0.1102	0.0536	0.0681	0.1089	0.0854
n	10	80	0.1072	0.2855	0.1864	0.0658	0.1418	0.0888	0.0744	0.1192	0.1254
o	15	80	0.0688	0.2218	0.0539	0.0653	0.1706	0.0732	0.0996	0.1301	0.2104
p	30	80	0.0708	0.2465	0.0635	0.0495	0.2406	0.0443	0.0890	0.1764	0.1319
q	5	100	0.1644	0.1374	0.1799	0.0522	0.1106	0.0512	0.0957	0.1134	0.1427
r	10	100	0.1333	0.2189	0.2659	0.0661	0.1414	0.0750	0.0674	0.1213	0.1100
s	15	100	0.0617	0.2121	0.0556	0.0622	0.1523	0.0859	0.0785	0.1330	0.1157
t	30	100	0.0805	0.2158	0.0705	0.0550	0.2218	0.0547	0.0798	0.1727	0.1357
avg.			0.0996	0.2761	0.1148	0.0675	0.1986	0.0780	0.0794	0.1568	0.1165

[a]MSE_{100} and MSE_{mid} are the MSE of the 1−100 and 201−400 sampling steps, respectively.

where κ, ρ, c_p, and L are the conductivity, the mass density, the specific heat capacity, and the unit mass latent heat, respectively, and I_0 and g represent the given intensity and the normalized profile. The crystallinity of the amorphous region after annealing, being molten and solidified, depends on the cooling rate assumed to be small enough to crystallize the solidified region completely. Then, the consequence is obtained by solving the crystallinity phase field equation, and the phase field variable φ is introduced to describe whether the material is liquid or solid.[47]

$$\Delta\alpha = -\alpha\left(1 - \frac{1}{2}\mathrm{erfc}\left(\log\left(\frac{\varphi_{rec}}{\varphi + \varepsilon\varphi_{rec}}\right)\right)\right) \tag{11}$$

where $\Delta\alpha$, φ_{rec}, and ε are the change of crystallinity per time interval, the parameter to manipulate which melting status starts to lose crystallinity information, and a scaling factor of φ_{rec} to avoid a zero denominator, respectively. Although the melting phase field equation for the melting laser model is more suitable in the case of our experiments, we have a problem with the convergence of the simulation results while using it. Then, the dopant diffusion equation can be solved by coupling it with the heat and phase equations.[47]

$$\frac{\partial C}{\partial t} = \nabla\cdot\left(D\frac{C_{eq}}{C_{eq0}}\nabla\left(C\frac{C_{eq0}}{C_{eq}}\right)\right) \tag{12}$$

$$\frac{C_{eq}}{C_{eq0}} = \exp\left(-\frac{\varphi E_{seg} + 16\varphi^2(1-\varphi)^2 E_{intf}}{kT}\right) \tag{13}$$

E_{seg} and E_{intf} are chemical potential energies of a solid state and an interface state corresponding to a liquid state. The sheet resistance formula is given by[47]

$$R_s = \frac{1}{\int_{pn_i}^{pn_{i+1}} q(\mu_n n + \mu_p p)dx}\Delta\alpha = -\alpha\left(1 - \frac{1}{2}\mathrm{erfc}\left(\log\left(\frac{f_{rec}}{f + \varepsilon f_{rec}}\right)\right)\right) \tag{14}$$

where electron and hole concentrations are calculated assuming charge neutrality at a temperature of 300 K, and μ is the mobility of electrons or holes defined by the formula below

$$\mu = \mu_{min} + \frac{\mu_{max} - \mu_{min}}{1 + \left(\frac{N}{N_r}\right)^{\alpha}} \tag{15}$$

where N is the background concentration, and μ_{min}, μ_{max}, N_r, and α are fitting parameters for the empirical formula.[49]

3. RESULTS AND DISCUSSION

Table 1 lists the mean square error (MSE) at the first 100 sampled data points (MSE_{100}), at the 200−400 sampled data points (MSE_{mid}), and over the entire sampling period (MSE). It is clearly seen in Table 1 that the initial advantage of using the hybridization of TCAD and ML is reflected in MSE_{100}. The first column in Table 1 lists the period that the base and meta-learners are retrained during sampling. The second column lists the number of epochs used in the retraining. Figure 2 is the plotting of the MSE of the test set versus the number of trained samples during the sampling steps. The sampling in this work aims at promoting the construction of an accurate ML model using fewer data points. The test set is the data points that have not been sampled. From Figure 2, it is observed that the test set MSE trend exhibits a different nature if we compare three cases, that is, the pure machine-learning-based sampling, the hybridized sampling using TCAD and ML, and pure TCAD-based sampling. The hybridization of TCAD

https://doi.org/10.1021/acsomega.2c06000
ACS Omega 2023, 8, 737−746

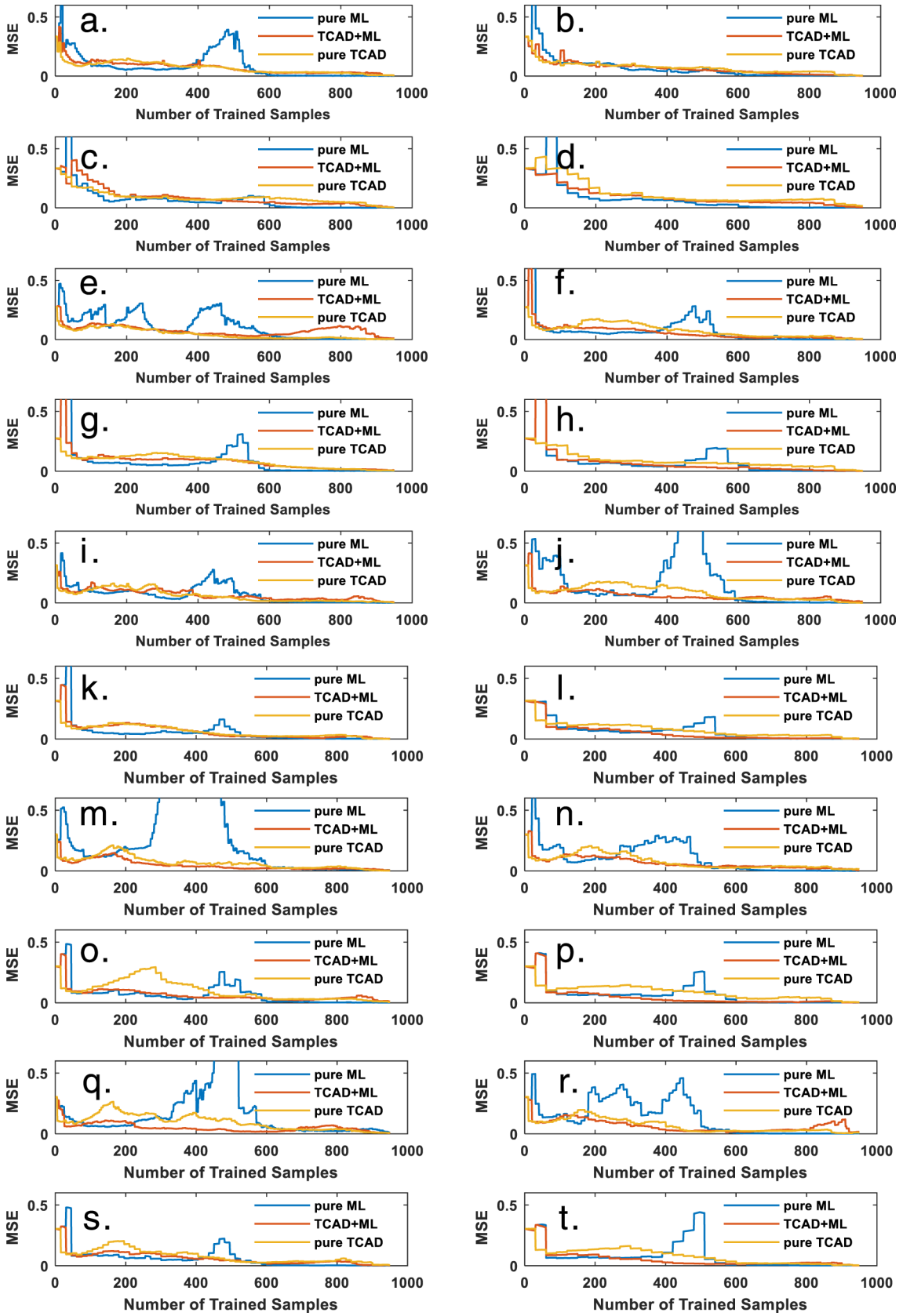


Figure 2. Mean squared error vs the trained sample number for pure neural-network-based sampling (blue lines), meta-learned, TCAD-assisted ML sampling (red lines), and pure TCAD-based sampling scheme (yellow lines) for the cases (a−t) in Table 1.

and ML has been explained in detail in Section 2. Pure ML-based sampling is deciding the next sampling points based on the currently available ML model. Optimization is done to locate the most appropriate data points. In this work, we will

https://doi.org/10.1021/acsomega.2c06000
ACS Omega 2023, 8, 737−746

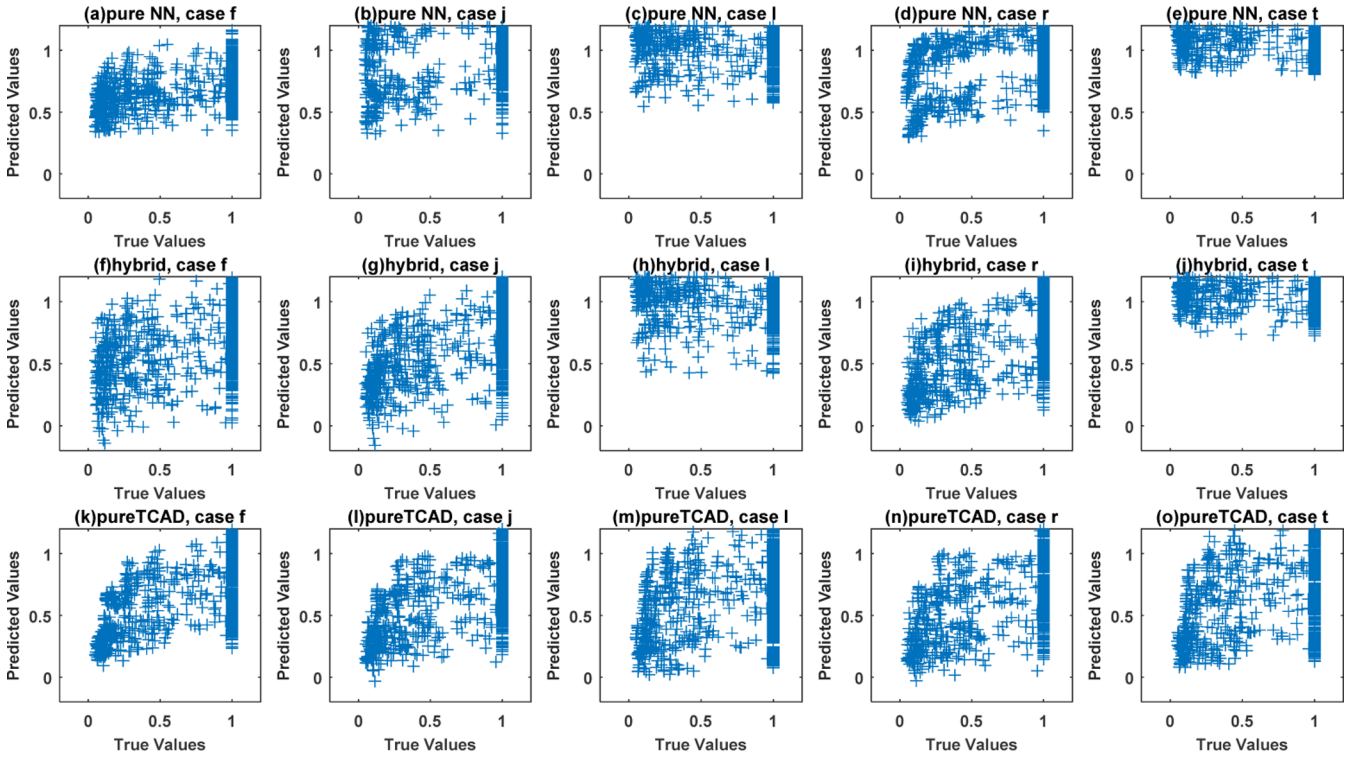


Figure 3. Test set prediction at the sampling step 50 for the different cases (f,j,l,r,t) in Table 1.

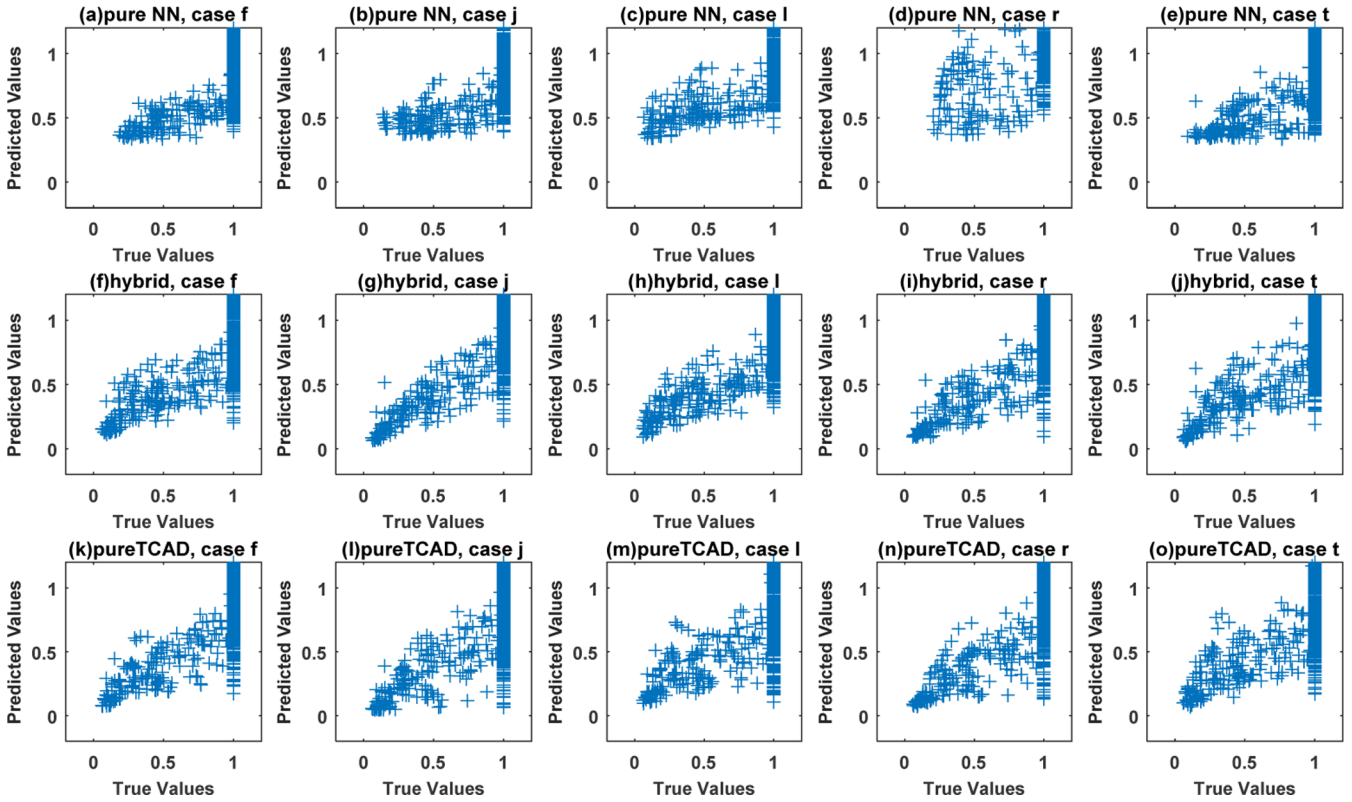


Figure 4. Test set prediction at the sampling step 300 for the different cases (f,j,l,r,t) in Table 1.

locate the data points exhibiting the largest deviation from the predicted-value average based on the partially trained network. Here, we assume prescribed discrete input process parameter values, while the same algorithm can be applied to a continuous value case. The only difference is that the optimization algorithm to locate the nest sampling point needs to be changed to a continuous variable version. It is seen from Figure 2 that, in most cases, the TCAD and hybrid algorithm exhibit better performance in the first portion of sampling. The first portion varies from case to case, and the

https://doi.org/10.1021/acsomega.2c06000
ACS Omega 2023, 8, 737−746

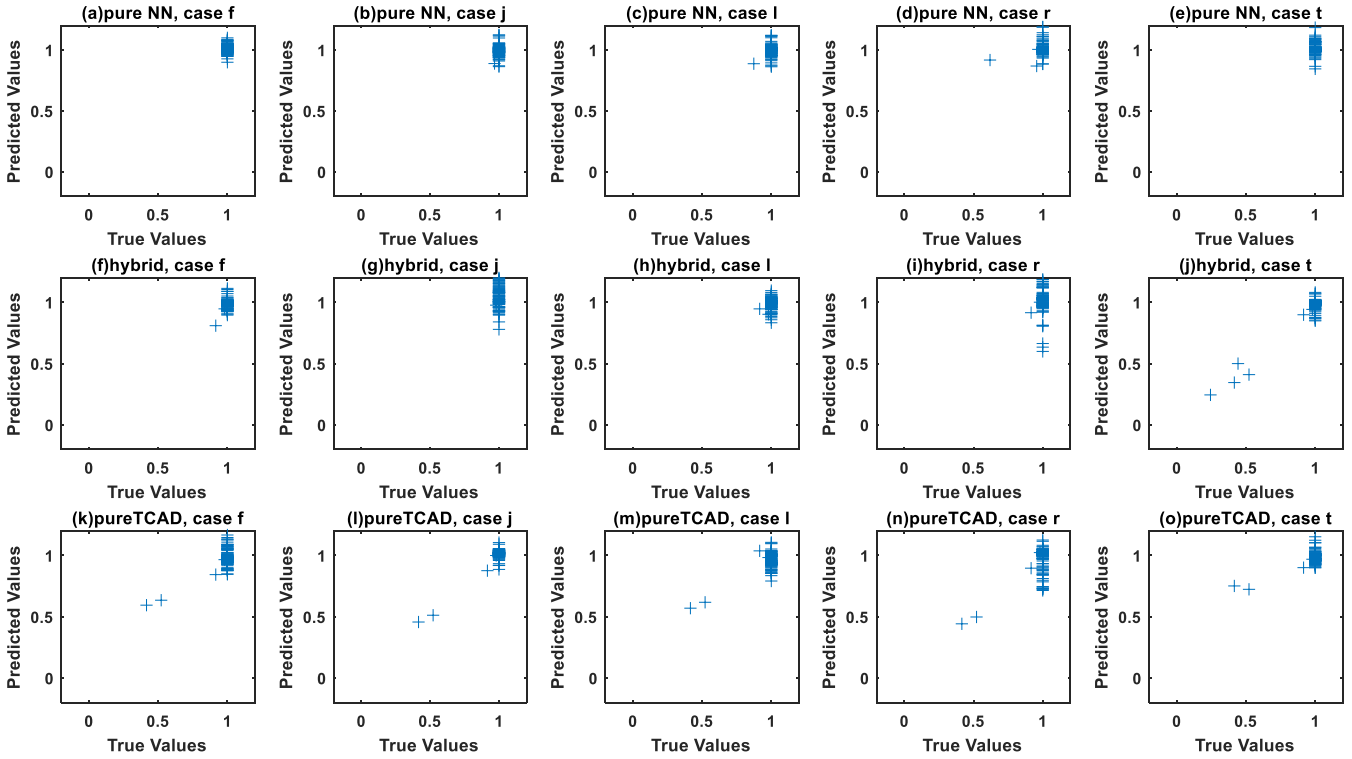


Figure 5. Test set prediction at the sampling step 950 for the different cases (f,j,l,r,t) in Table 1.

range falls into 30−100 steps in the 20 cases in this work. The reason for improved performance is that the neural network model has not been collecting enough amount of data to give reasonable predictions. This is the weakness of essentially all ML models, where a certain amount of data needs to be collected to fulfill reasonable prediction accuracy. In Figure 2, the TCAD-based sampling can become less efficient at the mid-stage, that is, 200−400 sampling steps, because the TCAD is based on analytical modeling where the discrepancy between theory and experiment always exists. It is quite impossible to use TCAD to model many subtle phenomena, and many processes and experiments in semiconductor manufacturing have not been properly described by TCAD models. Even for laser thermal annealing, whose analytical model is indeed explained by several equations in Section 2.2, the discrepancy from the experiment is still pronounced. At a later stage, at 400−600 sampling steps, it is observed that the ML model again shows some increased MSE at the test set. Fortunately, the meta-learner effectively prevents the hybrid model from following trends of pure ML-based sampling. It should be emphasized that the phenomenon regarding the relative strength of ML and TCAD models can vary case by case, except possibly for the initial weakness of the ML model. Therefore, it is important to have a meta-learner to guide the sampling in order to receive the benefit of hybridization. At a later stage (>900 data points) during the sampling procedure, the convergence of the three models in Figure 2 is observed.

It is worth mentioning that the base learner is periodically changing and improving its accuracy with new samples collected and retraining conducted. Nevertheless, reassigning the true labels for the classification of meta-learner is not required and actually should not be conducted for the previously sampled process parameters x. This is because the trained base learner has full accuracy at previous sampling points, and thus, reassigning the classification of true labels for the previously sampled data only leads to strong biases to the $x_{\text{next, ML}}$. Indeed, the gradually improved base learner can lead to changed true labels as far as the classification meta-learner is concerned, and in order to take into account this effect, a more sophisticated strategy such as a further partitioned training set during active learning is required to re-compare the strength of ML and TCAD prediction at all of the previously sampled data points. We do not implement such a complicated strategy in this study, and the effectiveness of the algorithm is already quite pronounced, which can be due to the fact that the current sampling is closer to the more recent sampling data in the sample space according to the strategy used in this study. Thus, the accuracy of the classification meta-learner at the more recent sampling step is more important.

Figures 3−5 further show the scatter plots of true versus predicted values for the test data set at the 50th, 300th, and 950th sampling steps for a pure neural network-based sampling, a meta-learned, TCAD-assisted ML sampling, and a pure TCAD-based sampling scheme. From Figure 3, it is observed that at the 50th step, the prediction on the test data set is not effective by pure ML, where the trend between the predicted values (y_{ML}) and the true values (y_{true}) has not established by the 50 sampled data points. On the other hand, the data trend has been seen in TCAD-assisted ML and pure TCAD, due to the effective guidance of theory in the early stage of sampling. Figure 4 shows the mid-stage sampling at the 300th sampling step. It is seen that pure ML is improved and is comparable to the pure TCAD sampling. Comparing TCAD-assisted ML and pure TCAD at the 300th sampling step, it can be observed that pure TCAD is inferior due to the inaccuracy of the pure theoretical framework, which is not corrected by the experimental data. Figure 5 shows the late-stage sampling at the 950th sampling step, and similar accurate predictions are achieved for all of the three algorithms on the remaining small amount of test data.

https://doi.org/10.1021/acsomega.2c06000
ACS Omega 2023, 8, 737−746

Meta-learning, in general, refers to using an algorithm to adjust the learning process, which can be (1) the network architecture, such as pruning, connection, and neuron number; (2) basic ML hyperparameters, such as learning rate, batch size, optimizers, and so forth; (3) data collection method such as the case in this work; and (4) other aspects that can affect the model training. Suppose hyperparameters are defined as any parameters that will not be tuned during model training. In that case, meta-learning can be defined as using another model, that is, meta-learner, to tune the hyperparameters. While in many contexts, hyperparameters are only referred to as the basic training parameters, such as batch size and learning rate, many more aspects can be regarded as hyperparameters, including network architectures. It should be clear that in this work, we try to use a meta-learner to control the behavior of sampling at each step of active learning.[50−52] The base learner is the network predicting the sheet resistance as a function of process parameters. Before the base learner learns the input−output relation, the way to decide the next sample between TCAD and ML models is tuned first based on the past sampling experience.

The way to utilize the past sampling experience to enhance the meta-learner and to facilitate the speed of constructing an accurate base learner has firstly count on deciding the input to the meta-learner. Clearly, the most straightforward strategy is to use the process parameters as the input and consider whether the TCAD model or the ML model is more accurate as the output, consisting of a classification problem. In this case, the input to the base learner and to the meta-learner is the same. The selection of the next sampling point is based on tracing the currently available ML model and the existing TCAD model and locating the maximal deviation. Then, using ML or TCAD model is based on the accuracy of the two models at various locations on the input sample space. Alternatively, long-short-term memory (LSTM) or time-series-based method can also be used to model the time evolution of sampling. In this case, the correlation between the input and output is related to the mapping between the process parameters and not only the classified accuracy of the ML and TCAD models but also the previous sampling sequence in the sense of LSTM. This means that more recent sampling affects the next sampling more than the less recent sampling does. We do not employ the time series analysis method. This is because while the sequence of sampling may have some effect on the model prediction accuracy, using LSTM requires repeated construction of imaginary sampling sequences, in the currently available training set, to supplement the real sampling sequence. While the effect of the sampling sequence and the effectiveness of constructing imaginary sampling sequences is unclear, time-series analysis techniques are not considered in this work.

Regarding the TCAD model, our modeling is based on commercially available TCAD software. The diffusion, heat, melting phase, and crystallinity equations lead to a coupled solution. The partial differential equation on time and space is solved; thus, we can see the evolution of dopant activation during annealing as a space function. The substrate temperature can be varied in our experimental setup, but the TCAD model can only be set at its ambient temperature, and this aspect leads to some inaccuracy. Essentially, the wafer top surface in the TCAD model is described by the Stefan−Boltzmann law, where the radiation of the thermal energy to the ambiance is calculated as the boundary condition. The wafer bottom surface is set as the thermal conduction boundary where the temperature difference between the wafer and the ambiance leads to the heat flow between the material contact. It is impossible to set two temperatures in the TCAD model; thus, we assume that the stage temperature and the air on the wafer's top surface temperature are the same. Assuming that the wafer thickness is thin compared to the temperature gradient from the stage to the sample top surface, the Stefan Boltzmann law still holds in the case of stage heating. Another model inaccuracy lies in the fact that pulsed laser condition can only be modeled as a continuous wave or a single or a few laser pulses. In this case, the effect of laser repetition rate and the nonlinear behavior of high-power laser pulses cannot be fully modeled. In our simulation, the repetition rate is modeled by adjusting the laser fluence and anneal time. On the other hand, the TMM model is sufficient to model planar multilayers, and thus, electromagnetic wave modeling is not required in our poly-Si/SiO_2/Si-substrate case.

4. CONCLUSIONS

Constructing an accurate prediction model efficiently and at all sampling stages is important and desired in the semiconductor industry. To achieve this goal, an efficient sampling scheme is required. TCAD-assisted sampling is proposed in this work. The approach supplements the deficiency of ML-based sampling during the initial data collection and the deficiency of inaccurate TCAD models in the mid-stage. Using laser thermal annealing as an example, we have demonstrated the effectiveness of the meta-learned sampling procedure. While the base learner is gradually improving its accuracy by active learning, the classification meta-learner effectively selects the more appropriate next sampling candidate from TCAD or ML suggestion, which will be incorporated into the base learner training. The early, that is, <100 steps, and mid-stage, that is, 200−400 steps, strength is observed by using TCAD-assisted ML sampling. The future improvement will further improve meta-learner to fully utilize the TCAD and ML sampling flexibly, verifying the effectiveness of the proposed algorithm by testing it on a variety of semiconductor processes and going beyond the multiplayer perceptron. We believe that the meta-learned, TCAD-assisted sampling prediction network will be beneficial in nearly all future semiconductor manufacturing experiments.

AUTHOR INFORMATION

Corresponding Author

Albert Shihchun Lin − *Institute of Electronics Engineering, National Yang Ming Chiao Tung University, Hsinchu City 30010 Taiwan, R.O.C;* orcid.org/0000-0001-6104-3360; Email: hdtd5746@gmail.com

Authors

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

Chung Yuan Chang − *Institute of Electronics Engineering, National Yang Ming Chiao Tung University, Hsinchu City 30010 Taiwan, R.O.C;* orcid.org/0000-0001-5725-9783

Yen-Wei Feng − *Institute of Electronics Engineering, National Yang Ming Chiao Tung University, Hsinchu City 30010 Taiwan, R.O.C;* orcid.org/0000-0003-0764-2092

https://doi.org/10.1021/acsomega.2c06000
ACS Omega 2023, 8, 737−746

ShihWei Chen − *Taiwan Semiconductor Research Institute, Hsinchu City 300091 Taiwan, R.O.C;* orcid.org/0000-0002-1769-8519
Chang-Hong Shen − *Taiwan Semiconductor Research Institute, Hsinchu City 300091 Taiwan, R.O.C*
Jia-Min Shieh − *Taiwan Semiconductor Research Institute, Hsinchu City 300091 Taiwan, R.O.C;* orcid.org/0000-0001-5266-1052

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c06000

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported in part by the Ministry of Science and Technology, Taiwan, under Grant MOST 111-2221-E-A49-186.

■ REFERENCES

(1) Wu, Z.; Chu, W.Sampling Strategy Analysis of Machine Learning Models for Energy Consumption Prediction. *2021 IEEE 9th International Conference on Smart Energy Grid Engineering (SEGE), 11−13 Aug. 2021*, 2021; pp 77−81.
(2) Verrelst, J.; Berger, K.; Rivera-Caicedo, J. P. Intelligent Sampling for Vegetation Nitrogen Mapping Based on Hybrid Machine Learning Algorithms. *IEEE Geosci. Rem. Sens. Lett.* **2021**, *18*, 2038−2042.
(3) Chu, W.; Ho, P. S.; Li, W. An Adaptive Machine Learning Method Based on Finite Element Analysis for Ultra Low-k Chip Package Design. *IEEE Trans. Compon. Packag. Manuf. Technol.* **2021**, *11*, 1435−1441.
(4) Ramezan, C. A.; Maxwell, T. A.; Maxwell, A. E. Evaluation of Sampling and Cross-Validation Tuning Strategies for Regional-Scale Machine Learning Classification. *Remote Sens.* **2019**, *11*, 185.
(5) ElRafey, A.; Wojtusiak, J. Recent advances in scaling-down sampling methods in machine learning. *WIREs Comput. Stat.* **2017**, *9*, No. e1414.
(6) Homeyer, A.; Schenk, A.; Dahmen, U.; Dirsch, O.; Huang, H.; Hahn, H. K. A comparison of sampling strategies for histological image analysis. *J. Pathol. Inf.* **2011**, *2*, S11.
(7) Lecarpentier, L.; Mili, A.; Decaunes, J.; Le Gratiet, B.New approach for APC and measurement sampler interaction in a complex process mix logic fab. *Metrology, Inspection, and Process Control for Microlithography XXXIII*; SPIE: 2019; Vol. *10959*.
(8) McLoone, S.; Johnston, A.; Susto, G. A. A Methodology for Efficient Dynamic Spatial Sampling and Reconstruction of Wafer Profiles. *IEEE Trans. Autom. Sci. Eng.* **2018**, *15*, 1692−1703.
(9) Susto, G. A.A dynamic sampling strategy based on confidence level of virtual metrology predictions. *2017 28th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC), 15−18 May 2017*, 2017; pp 78−83.
(10) Sahnoun, M. h.; Bettayeb, B.; Bassetto, S.-J.; Tollenaere, M. Simulation-based optimization of sampling plans to reduce inspections while mastering the risk exposure in semiconductor manufacturing. *J. Intell. Manuf.* **2016**, *27*, 1335−1349.
(11) Nduhura-Munga, J.; Rodriguez-Verjan, G.; Dauzere-Peres, S.; Yugma, C.; Vialletelle, P.; Pinaton, J. A Literature Review on Sampling Techniques in Semiconductor Manufacturing. *IEEE Trans. Semicond. Manuf.* **2013**, *26*, 188−195.
(12) Prakash, P.; Honari, B.; Johnston, A.; McLoone, S. F.Optimal wafer site selection using forward selection component analysis. *2012 SEMI Advanced Semiconductor Manufacturing Conference, 15−17 May 2012*, 2012; pp 91−96.
(13) Hong, S. J.; Lim, W. Y.; Cheong, T.; May, G. S. Fault Detection and Classification in Plasma Etch Equipment for Semiconductor Manufacturing e-Diagnostics. *IEEE Trans. Semicond. Manuf.* **2012**, *25*, 83−93.
(14) Asano, M.; Ikeda, T. Sampling plan optimization for critical dimension metrology. *J. Nanolithogr. MEMS, MOEMS* **2006**, *5*, 033008.
(15) Wan, J.; Honari, B.; McLoone, S.A dynamic sampling methodology for plasma etch processes using Gaussian process regression. *2013 XXIV International Conference on Information, Communication and Automation Technologies (ICAT)*; IEEE: Sarajevo, Bosnia and Herzegovina, 2013.
(16) Batalov, R. I.; Nuzhdin, V. I.; Valeev, V. F.; Vorobev, V. V.; Osin, Y. N.; Ivlev, G. D.; Stepanov, A. L. Pulsed laser annealing of high-dose Ag+-ion implanted Si layer. *J. Phys. D: Appl. Phys.* **2017**, *51*, 015109.
(17) Boyd, I. W.; Wilson, J. I. B. Laser annealing for semiconductor devices. *Nature* **1980**, *287*, 278.
(18) Robinson, A. L. Laser Annealing: Processing Semiconductors Without a Furnace. *Science* **1978**, *201*, 333−335.
(19) Moparthi, S.; Tiwari, P. K.; Saramekala, G. K. Sensitivity Analysis of Silicon Nanotube FET (Si NTFET) with TCAD Assisted Machine Learning. *Silicon* **2022**, *14*, 9021.
(20) 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.
(21) 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.
(22) Ren, H.; Stewart, R.; Song, J.; Kuleshov, V.; Ermon, S. Learning with Weak Supervision from Physics and Data-Driven Constraints. *AI Magzine* **2018**, *39*, 27−38.
(23) Zhao, P.; Liu, Y. Physics Informed Deep Reinforcement Learning for Aircraft Conflict Resolution. *IEEE Trans. Intell. Transport. Syst.* **2022**, *23*, 8288−8301.
(24) Pratik, S.; Liu, P.-N.; Ota, J.; Tu, Y.-L.; Lai, G.-W.; Ho, Y.-W.; Yang, Z.-K.; Rawat, T. S.; Lin, A. S. Mapping Oxidation and Wafer Cleaning to Device Characteristics Using Physics-Assisted Machine Learning. *ACS Omega* **2022**, *7*, 933−946.
(25) Parashar, P.; Chen, C. H.; Akbar, C.; Fu, S. M.; Rawat, T. S.; Pratik, S.; Butola, R.; Chen, S. H.; Lin, A. S. Analytics-statistics mixed training and its fitness to semisupervised manufacturing. *PLoS One* **2019**, *14*, No. e0220607.
(26) 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. *Comput. Stat. Data Anal.* **2019**, *132*, 126−142.
(27) Liu, N.; Wu, Z.; Li, G.; Liu, X.; Wang, Y.; Zhang, L. MAIC: Metalearning-Based Adaptive In-Field Calibration for IoT Air Quality Monitoring System. *IEEE Internet Things J.* **2022**, *9*, 15928−15941.
(28) Lee, J.; Yang, C. Deep neural network and meta-learning-based reactive sputtering with small data sample counts. *J. Manuf. Syst.* **2022**, *62*, 703−717.
(29) Yu, N.; Chen, H.; Xu, Q.; Hasan, M. M.Wafer map defect recognition with few shot learning based on hybrid self-attention mechanism and prototype network. *2021 China Automation Congress (CAC)*; IEEE, 2021; pp 4128−4134.
(30) Maudsley, D. B.*A Theory of Meta-Learning and Principles of Facilitation: An Organismic Perspective*; University of Toronto, 1979.
(31) Yu, H.; Kim, S.In Passive Sampling for Regression. *2010 IEEE International Conference on Data Mining, 13−17 Dec. 2010*, 2010; pp 1151−1156.
(32) Settles, B.Active learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*; Morgan & Claypool Publishers, 2012; Vol. *6*, pp 1−114.
(33) Raj, A.; Bach, F.In Convergence of uncertainty sampling for active learning, *International Conference on Machine Learning, PMLR*: 2022; pp 18310−18331.
(34) Nguyen, V.-L.; Shaker, M. H.; Hüllermeier, E. How to measure uncertainty in uncertainty sampling for active learning. *Mach. Learn.* **2022**, *111*, 89−122.

(35) Dasgupta, S. Two faces of active learning. *Theor. Comput. Sci.* **2011**, *412*, 1767−1781.
(36) Freund, Y.; Seung, H. S.; Shamir, E.; Tishby, N. Selective Sampling Using the Query by Committee Algorithm. *Mach. Learn.* **1997**, *28*, 133−168.
(37) Cohn, D.; Atlas, L.; Ladner, R. Improving Generalization with Active Learning. *Mach. Learn.* **1994**, *15*, 201−221.
(38) Schein, A. I.; Ungar, L. H. Active learning for logistic regression: an evaluation. *Mach. Learn.* **2007**, *68*, 235−265.
(39) Hoi, S. C. H.; Jin, R.; Lyu, M. R.Large-scale text categorization by batch mode active learning. *Proceedings of the 15th International Conference on World Wide Web*; Association for Computing Machinery: Edinburgh, Scotland, 2006; pp 633−642.
(40) MacKay, D. J. C. Information-Based Objective Functions for Active Data Selection. *Neural Comput.* **1992**, *4*, 590−604.
(41) Rawat, T.; Chung, C.-Y.; Chen, S.-W.; Lin, A.Reinforcement Learning based Intelligent Semiconductor Manufacturing Applied to Laser Annealing. **2022**, Engrxiv PrePrint DOI: 10.31224/2449.
(42) Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; et al.TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. **2015**, arXiv:1603.04467
(43) Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; Kern, R.; Picus, M.; Hoyer, S.; van Kerkwijk, M. H.; Brett, M.; Haldane, A.; del Río, J. F.; 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.
(44) McKinney, W.Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference (SciPy 2010)*, 2010.
(45) Rossum, G. V.; Drake, F. L.*Python 3 Reference Manual*, CreateSpace, 2009.
(46) Jones, E.; Oliphant, T.; Peterson, P.*SciPy: Open Source Scientific Tools for Python*, Comput. Program Opensource Library; Brigham Young Univ.: UT, USA, 2001.
(47) *Sentaurus Process User Guide*; Version L-2016.03, Synopsys, Inc., 690 East Middlefield Road Mountain View, CA 94043, 2016.
(48) *MathsWorks Matlab, 11.16.2.1 (R2019b)*; The MathWorksInc., 2019.
(49) Masetti, G.; Severi, M.; Solmi, S. Modeling of carrier mobility against carrier concentration in arsenic-, phosphorus-, and boron-doped silicon. *IEEE Trans. Electron Devices* **1983**, *30*, 764−769.
(50) Shim, J.; Kang, S.; Cho, S. Active Learning of Convolutional Neural Network for Cost-Effective Wafer Map Pattern Classification. *IEEE Trans. Semicond. Manuf.* **2020**, *33*, 258−266.
(51) Lin, Y.; Alawieh, M. B.; Ye, W.; Pan, D. Z.Machine Learning for Yield Learning and Optimization. *IEEE International Test Conference (ITC)*, 2018.
(52) Zhuo, C.; Agarwal, K.; Blaauw, D.; Sylvester, D.Active learning framework for post-silicon variation extraction and test cost reduction. *2010 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2010.

https://doi.org/10.1021/acsomega.2c06000
ACS Omega 2023, 8, 737−746