TCAD Augmented Machine Learning for Semiconductor Device Failure Troubleshooting and Reverse Engineering

Y. S. Bankapalli and H. Y. Wong
Electrical Engineering, San Jose State University, CA, USA
hiuyung.wong@sjsu.edu

Abstract – Motivation and Achievement

Semiconductor device failure troubleshooting and device reverse engineering require expensive analyses such as SEM and TEM [1]. Machine learning (ML) has been used widely in the manufacturing process to enable early discovery of defects [2]. However, the authors are not aware of any extensive application of ML to analyze defects based on finished product electrical characteristics, such as Current-Voltage (IV) and Capacitance-Voltage (CV) curves, where defects include epi thickness and doping level variations. This is because, for matured process with high yield, the number of defective dies is limited, while for nascent process with low yield, the number of dies produced are limited. As result, it is difficult to obtain enough defective IV curves for accurate machine learning.

Using TCAD, in principle, a large number of IV's can be generated by changing the layer thicknesses (to model epi layer variation) and doping levels, and by including various defective models (such as trap assisted tunneling at various spatial location). ML can then be used to generate model to accurately correlate IV and CV curves to defects. Based on the trained model, one can rapidly narrow down the possible cause of an abnormal IV curve and, if necessary, perform further failure analysis (e.g. cutting TEM at the most probably failure spot predicted by ML). The same reasoning applies well in device reverse engineering.

In this paper, we demonstrate this idea by studying the relationship between 1-D PIN diode epi layer thickness variations (a type of fabrication defect) and its forward and reverse IV curves. Various machine learning models are tested. It is found that 1) data processing before ML is critical to obtain accurate results, 2) linear regression gives the best prediction and is better than Multi-Layer Perceptron (MLP), 3) the model is able to predict structure with thickness out of the range of training data set and 4) the full process (TCAD simulation and ML) can be completed in less than 2 days with 1 core. It is expected that similar throughput can be achieved in 3D cases by using computing farm with thousands of cores, which is commonly available to the industry.

TCAD Simulations

Figure 1 inset shows the structure simulated. About 2000 1D PIN Diode structures are created using SProcess [3] with n+/i/p+ thicknesses being varied independently within the range given in Figure 1. Figure 2 shows the scattering plots of n+/i/p+ thicknesses, which are uniform and independent. SDevice is then used to simulate the IV characteristics [4]. Essential physics models are turned on, including Fermi-Dirac statistic, High Field saturation model, and non-local Band to Band tunneling [4]. Moreover, 80-bit

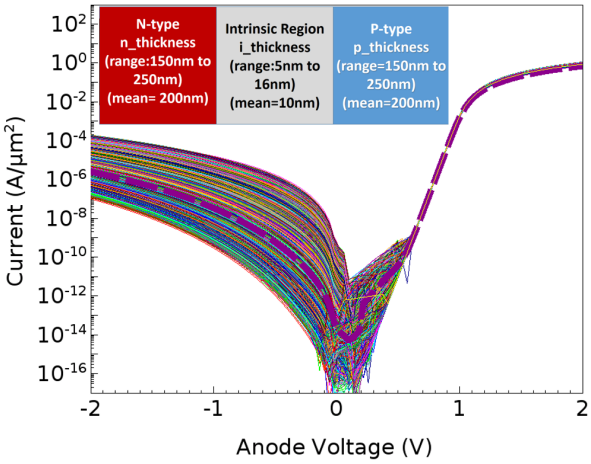

Figure 1: IV's of the 2000 devices simulated. The thick pink dash line is the IV of nominal device (200nm/10nm/200nm). Both n+ and p+ doping are $10^{20}cm^{-3}$.

ExtendedPrecision is used. Poisson, electron and hole continuity equations are solved self-consistently to produce the curves in Fig. 1.

Machine Learning

Scikit-learn library is used for ML. Four types of algorithms were tested, namely, linear regression (LR), decision tress (DT), random forest (RF) and Multi-Layer Perceptron (MLP) Regressor. 80% of the data (~1600) are used for training and 20% of the data (~400) are used for validation. The input is the IV curve (102 current values for V=-2V to 2V) and the output are i_thickness and n_thickness. Various parts of the IV curve are used for training, namely, I(V=-2V), I(V=-2V to 0V), I(V=0V to 2V) and I(V=-2V to 2V).

The first attempt to train with the raw data was not successful. This is because the current changes orders of magnitude for various thicknesses in reverse bias. As showed in Fig. 3, the model fails to predict large i_thickness (prediction is capped at about 12nm) because reverse current (I) is indistinguishable numerically in the raw form for large

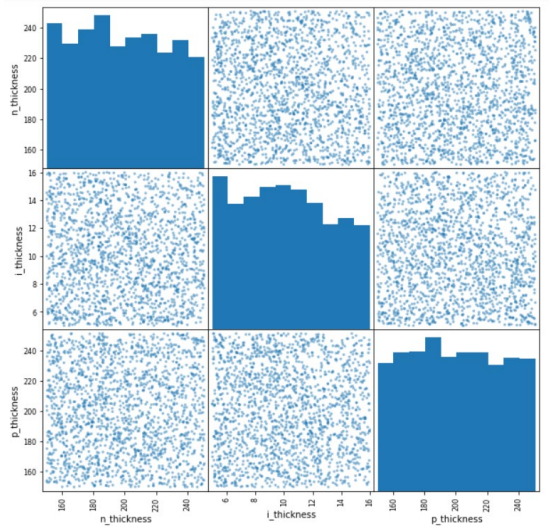

Figure 2: Scattering plot of n+/i/p+ thicknesses showing their frequencies and correlations.

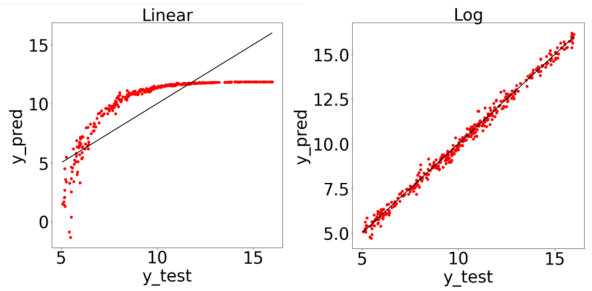

Figure 3: Prediction of i_thickness using linear regression model trained with raw current data (I, left) and processed current data (log(I), right) at -2V.

i_thickness. However, if log(I) is used, it gives good prediction. Therefore, in all trainings, log(I) is used.

Table 1 shows the i_thickness and n_thickness prediction Mean Squared Error (MSE) of various machines trained by different data ranges and algorithms. The learning can be summarized as:

1) DT is a not a suitable algorithm as it often overfits (training MSE = 0, with large prediction MSE)
2) LR performs the best with low training and prediction MSE for both i_thickness and n_thickness
3) MLP performs similar to LR for i_thickness but fails with n_thickness
4) Wide voltage range (-2V to 2V) gives the most accurate results. However, depends on the problem of interest, reduced voltage range gives similar results and simulation time can be substantially reduced. For example, by using the current at -2V, high accuracy of i_thickness can be obtained already because i_thickness influences the BTBT current strongly.
5) It is important to perform the simulation in a regime where the relevant physics is captured. For example, reverse current is insensitive to n_thickness. Therefore, bad result is obtained if data is only available between -2V to 0V. Positive bias simulation is required for n_thickness as forward neutral region potential drop correlates strongly to n_thickness. (Figure 4)

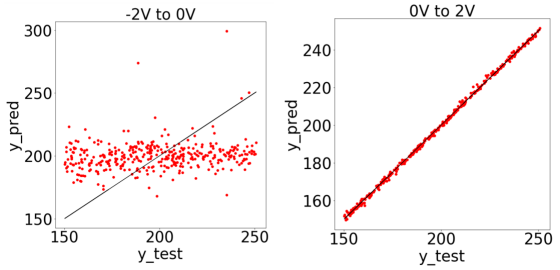

Figure 4: n-thickness prediction by LR machines trained by data from -2V to 0V (left) and data from 0V to 2V (right).

	Data Range Used	MLP	LR	DT	RF
i_thickness	"-2V"	0.06/0.06	0.06/0.06	0.03/0.08	0.01/0.08
	"-2V to 0V"	0.26/0.22	0.04/0.05	0.02/0.06	0.01/0.05
	"0V to 2V"	0.86/0.88	0.09/0.09	0.00/0.29	0.03/0.18
	"-2V to 2V"	0.05/0.05	0.03/0.03	0.01/0.04	0.01/0.04
n-thickness	"-2V"	847/796	846/795	0.00/1741	163/1248
	"-2V to 0V"	1043/984	761/811	0.00/1456	114/751
	"0V to 2V"	514/434	0.96/0.89	0.00/294	22/162
	"-2V to 2V"	473/407	0.80/0.86	86.91/236	22/163

Table 1: Training and predction Mean Squared Errors (MSE) of i_thickness and n_thickness by machines trained by various data range and algorithms. The numbers are format in "training MSE/ prediction MSE".

Prospect of 3D TCAD Simulation with ML

The 1D PIN diode has about 300 mesh points. The simulation was performed in Intel Xeon E5-2603 with 1 core used. The total simulation time of each simulation (process and device) is about 90 seconds. As a result, it takes about 2 days to complete data generation. A typical realistic 3D FinFET IV simulation is between 1 hours to 6 hours (process + device) [5]. If computing farm with thousands of CPU are available, we anticipate similar study can be completed in <1 day for realistic 3D FinFET structure. To reduce simulation time, one can reduce the number of training data point or/and the range of defect (e.g. i_thickness) variation. Fig. 5 show that even with 50 (or 200) data point, instead of 1600, MLP (or LR) is still very accurate. Moreover, LR can predict accurately 50% wider range of i_thickness than the training data. These make 3D TCAD augmented ML for defect trouble-shooting more feasible.

Conclusions

Using PIN diode with intrinsic layer and n+ layer thickness variations (a type of process defect), we demonstrated that TCAD can be used to generate sufficient data to train machine to identify the "defect value" rapidly based on IV curves. The task can be completed within 2 days. We anticipate that by using computing farm with thousands of cores, such scheme can be implemented for more realistic 3D simulations, especially in certain algorithm, only number of training data as low as 50 is needed. Such TCAD augmented ML can expedite defect trouble shooting and reverse engineering of semiconductor devices.

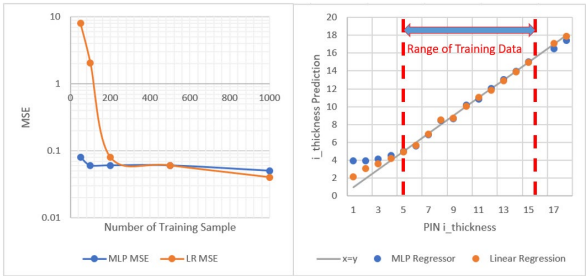

Figure 5: Prediction MSE as a function of the number of training samples (Left). Prediction of i_thickness as a function of PIN i_thickness (Right).

References

[1] R. Torrance and D. James, "Reverse Engineering in the Semiconductor Industry," 2007 IEEE Custom Integrated Circuits Conference, San Jose, CA, 2007, pp. 429-436. doi: 10.1109/CICC.2007.4405767

[2] G. A. Susto, M. T. and A. Beghi, “Anomaly Detection Approaches for Semiconductor Manufacturing”, Procedia Manufacturing 11 (2017) 2018 – 2024.

[3] Sentaurus™ Process User Guide Version O-2018.06, June 2018.

[4] Sentaurus™ Device User Guide Version O-2018.06, June 2018.

[5] H. Y. Wong, D. Dolgos, L. Smith and R. V. Mickevicius, “Modified Hurkx Band-to-Band-Tunneling Model for Accurate and Robust TCAD Simulations”, submitted to Mircoelectronics Reliabily.