Machine Learning in TCAD for Semiconductor Fabrication Failure Troubleshooting

by YOGESWARA SARAT BANKAPALLI

ABSTRACT:

The semiconductor manufacturing defects heavily impact the productivity in the industry. Identifying the defects manually is time-consuming. The root cause analysis of defects is very expensive. In the industry, machine learning is used to correlate failures to defects, but the data are limited. A large amount of data is generated by turning on/off fundamental defect models in TCAD which is impossible experimentally.

In this project we made the predictions of the underlying physical cause for the abnormality in Current-Voltage and Capacitance-Voltage curves. Various Machine Learning models are trained to understand the correlations between the inputs (I-V and/or C-V curves) and outputs (defects). After training, the machine automatically makes the predictions for the root cause of any abnormality in curves. To the best of our knowledge, this has not been demonstrated yet. This approach will reduce the technology development time and time to market for new product significantly.

Table of Contents

1. INTRODUCTION:

In semiconductor manufacturing industry, the defective devices cause the yield loss. The yield loss causes negative effects on the overall fabrication system performance. The yield loss heavily impacts the development cost and time to market of new technologies. Therefore, the identification of the root cause of the yield loss is very important in the semiconductor fabrication. The defective devices generally have abnormal properties in the IV (Current-Voltage) [1] and CV (Capacitance-Voltage) [2] curves. To identify the root cause of the abnormal properties, a physical analysis such as SEM (Scanning Electron Microscopy) and TEM (Transmission Electron Microscopy) [3] is required. However, the physical analysis is time-consuming, labor-intensive and expensive. In addition, the accurate prediction of the yield is critical.

An alternative approach for the detection of the defects is to use machine learning. Machine Learning is widely used in the semiconductor fabrication industry for the early detection of the defective devices. Machine Learning is a reliable approach for analyzing the defects based of the electrical characteristics (IV and CV curves) of the finished products. However, Machine Learning needs a large amount of data for accurate and reliable predictions. The required amount of data of the defective devices (defect types vs. IV and CV) is not available experimentally.

2. METHODOLOGY:

Technology Computer Aided Design (TCAD) [4] is used to generate a large amount of data. TCAD is usually used for development and optimization of semiconductor technologies. TCAD simulations can solve fundamental, physical, partial differential equations, etc. TCAD simulations are widely used in the semiconductor industry because of its intensive physical approach. This intensive physical approach of the TCAD simulations improves the predictive accuracy of the simulations. TCAD generates a large amount of data with different doping concentrations and thickness of the n+ region, Intrisic layer and p+ region, and different defective models.

Different Machine Learning models will be trained in such a way that, the trained models understand the correlations between the inputs (IV, CV) and outputs (defects). The trained models can filter the possibilities of the abnormal curves and thus improve the accuracy of the prediction. Based on the accuracy of the prediction, the model that gives the best accuracy will be selected for further analysis.

3. TCAD SIMULATIONS:

TCAD performs process simulation (SProcess) [5] and device simulation (SDevice) [6]. In process simulation, general processing steps such as oxidation, etching, ion implantation, etc., are performed. SProcess is one of the tools used for process simulation. The output of process simulation can be used for device simulation. SDevice is a tool used for device simulation. The illustration of the IV curves of 2000 devices (1D PIN Diode Structures) generated using SProcess is shown in Figure 1. Each curve has different n+/i/p+ thicknesses within a given range of uniform distribution. The doping concentrations of n+ and p+ are 10^{20} cm^{-3}. After the generation of the IV curves using SProcess, SDevice is then used for simulation of the characteristics of the IV curves. The data of the IV curves is exported as a csv (comma separated values) file. This comma separated values data is used for training the machine learning models.

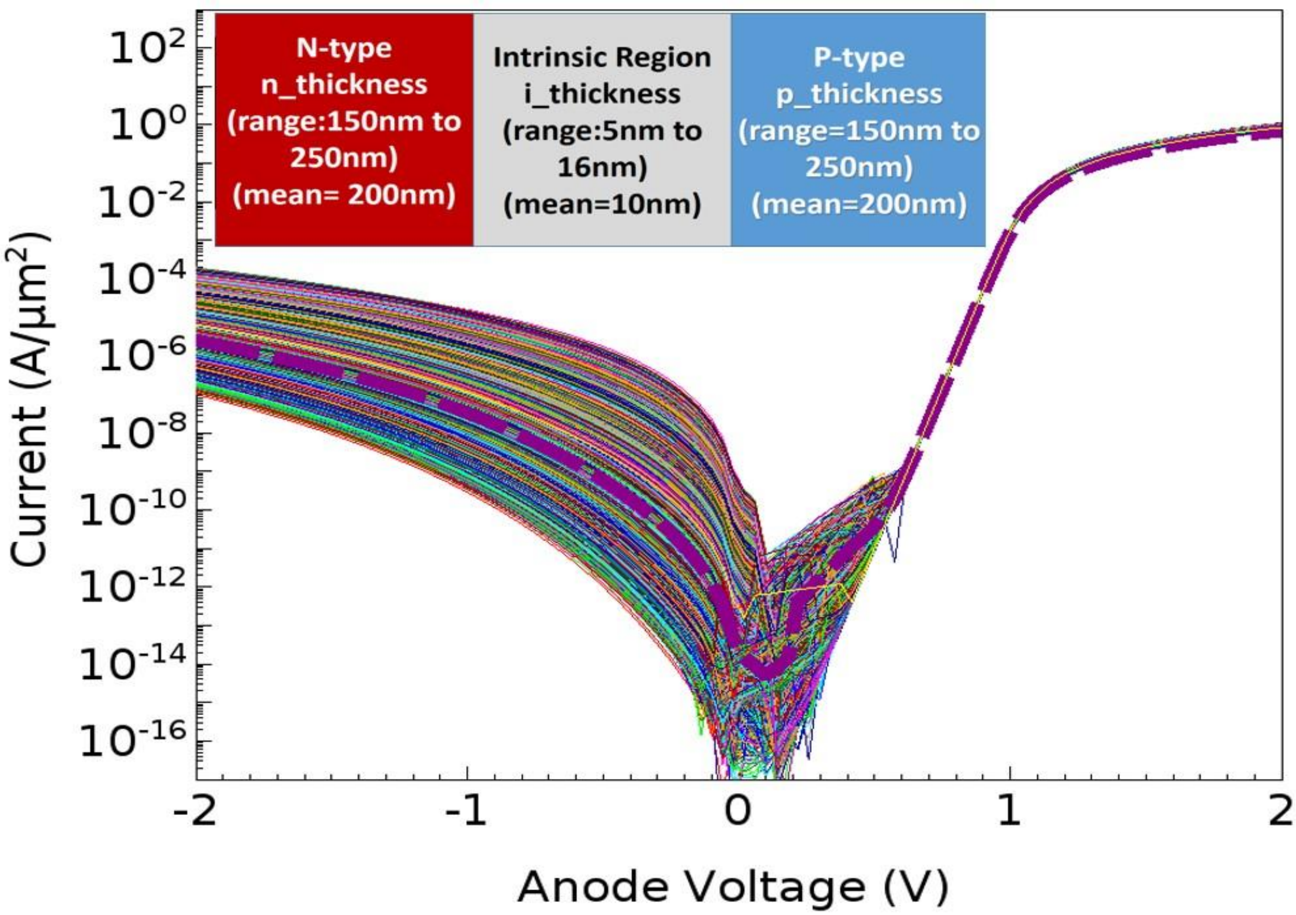

Figure 1: Current-Voltage (IV) curves of 2000 devices

4. DATASET:

4.1 Dataset Description:

The data is obtained as csv file from the TCAD Simulations. The dataset is a supervised learning dataset with regression problem. There are 2000 datapoints (2000 devices) in the dataset. The Input features are the current values (102 current values for -2V to +2V) of each IV curve. The output is the thickness values of n+ (N-type), p+ (P-type) and i (Intrinsic Region). However, the thickness of N-type region and P-type region are symmetrical. Therefore, there is no need to separately include the thickness of P-type region separately. The input features and the output are continuous values. There are no missing values in the dataset. However, there are few outliers in the dataset. The presence of outliers in the dataset is heavily affecting the predictions. Therefore, the outliers are removed from the dataset for good results. Initially, the data is divided into four different sets for training purpose, such as current values only at -2V, current values in between -2V to 0V, current values in between 0V to +2V, and current values in between -2V to +2V which is the entire dataset. Each set is split into training set and test set. 80% of the data (approximately 1600 datapoints) is used for training and 20% of the data (approximately 400 datapoints) is used for testing.

4.2 Data Visualization:

Data Visualization is one of the most important steps in dealing with large amount of data. The data visualization is an easy way to identify the outliers in the dataset. The data can be visualized in different ways such as box plots, scatter matrix, scatterplot, etc. The scatter matrix gives the pairwise scatter plot of the given features. The scatter matrix not only helps in identifying the outliers but also helps to identify the correlations between the given variables. The scatter matrix determines whether the given variables are positively correlated or negatively correlated. Figure 2 shows the scatter plot of the thicknesses of n+ region, intrinsic region and p+ region. The observations from the scatter matrix in figure 2 say that there is no correlation between n+ region, intrinsic region and p+ region thicknesses. The thicknesses of n+, intrinsic and p+ thicknesses are independent and are uniformly distributed within the given range.

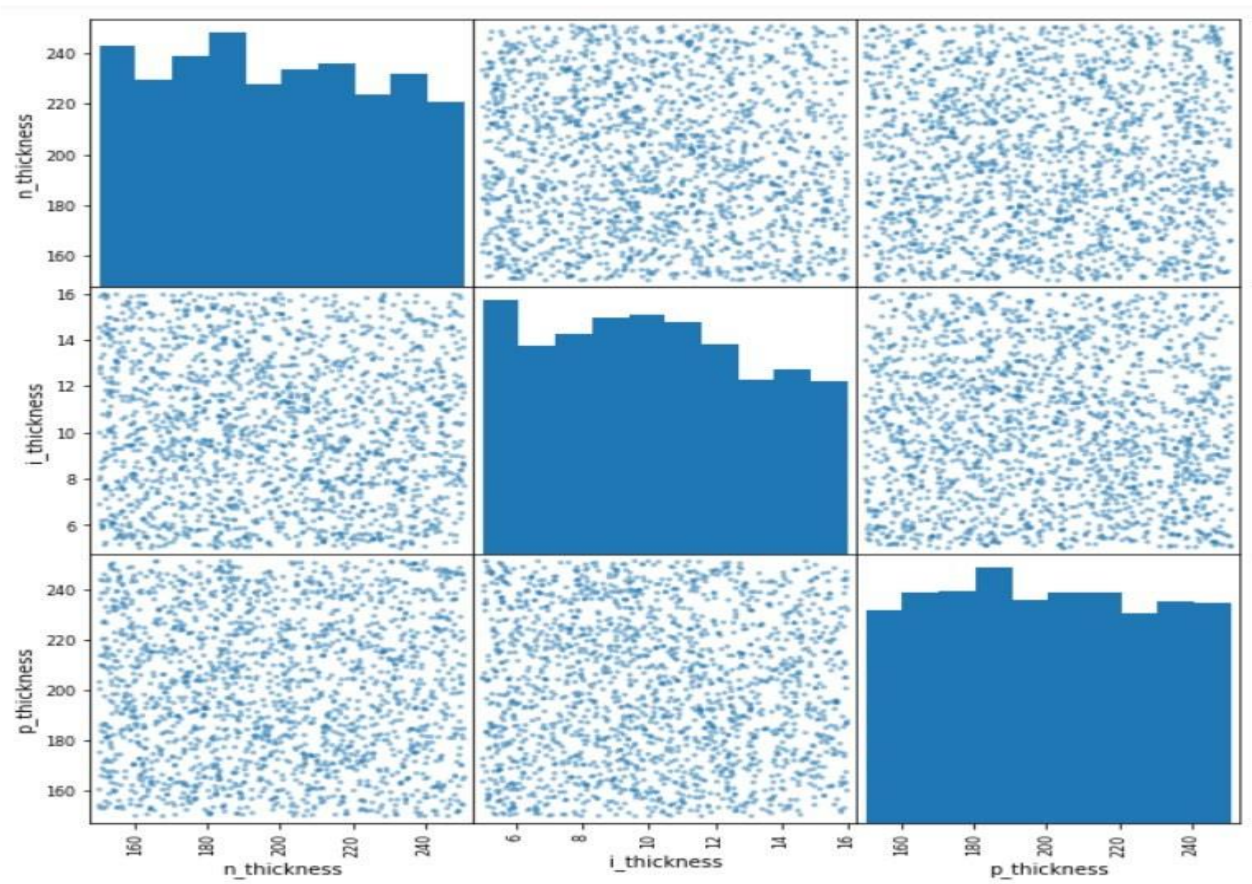

Figure 2: Scatter matrix between n+, i and p+ thicknesses.

5. MACHINE LEARNING:

The libraries used for this project are as follows: Scikit-Learn, pandas, numpy, matplotlib, subprocess. The data is trained using three machine learning models and one neural network model. The machine learning models are Linear Regression (LR), Decision Tree Regressor (DT) and Random Forest (RF). The neural network model is Multi-layer Perceptron (MLP) Regressor. All the four algorithms are tested on all the four different sets of data.

Initially, all the four algorithms are tested with the raw data (without any transformation). However, the attempt with the raw data is a failure, as the orders of magnitude of the thicknesses in the reverse bias is affected by the current. Therefore, the input data is transformed using the logarithmic scale. After the logarithmic transformation, the observed predictions are pretty good as compared to those with the raw data (without transformation) Therefore, the log-transformed data is used for further analysis. Figure 3 illustrates the comparison between predictions obtained with raw data and log-transformed data.

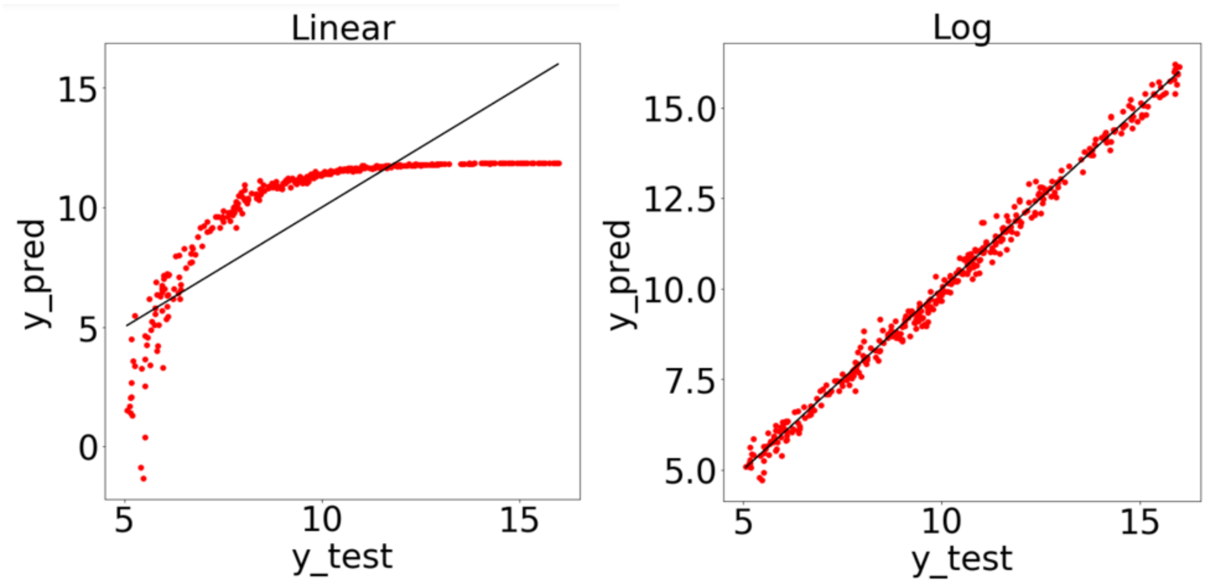

Figure 3. Comparison between the predictions of raw data and log-transformed data

5.1 Observations:

The four algorithms are applied on all the four sets of the log-transformed data. The observations in each set shown in the below table.

	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: i and n+ thicknesses MSE (training vs Test MSE) of all the four models for different data ranges

According to the observations, Linear Regression is performing well in predicting both n_thickness and i_thickness when compared to the other models. Linear Regression predictions have low Mean Square Error values on the training set as well as test set for both i-thickness and n-thickness. MLP is performing well only for i-thickness. The Mean Square Error values for MLP predictions for n-thickness on both training and test cases is too high. The Decision Tree algorithm also performs well for i-thickness. However, it overfits for n-thickness with 0 Mean Square Error values for training and very high MSE on the test case.

5.1.1 Training on I (-2V) dataset:

Linear Regression predictions for both n-type and intrinsic thicknesses are show in figure 4 when the training is performed only on current values at -2V. Clearly, Linear Regression is not able to perform well for n-type region when only the current values at -2V is used for training.

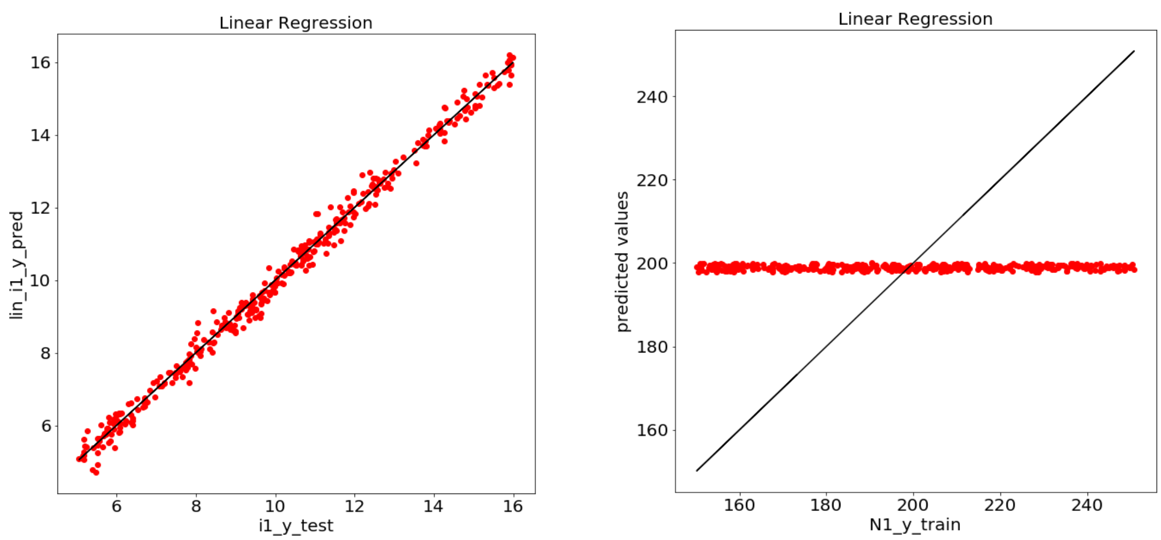

Figure 4: Linear Regression predictions on intrinsic (left) and n-type (right) thicknesses for current values at -2V

5.1.2 Training on I (-2V to 0V) dataset:

Figure 5 illustrates the Linear Regression predictions for both n-type and intrinsic when the training is performed only on current values in the region between -2V to 0V. The predictions are somewhat better compared to the prediction in -2V region. However, the performance of the Linear Regression is still not good for n-type region when only the current values at -2V to 0V is used for training. By observing the predictions when using only the current values at -2V and the predictions when using the current values between -2V to 0V, the prediction accuracy is slightly improved with the increase in the voltage range. To justify this statement, the data is further divided into different split, such as I (-2V to -1V) and I (-1V to 0V). The Machine Learning model is then applied on these splits. The obtained results show that the prediction accuracy gets slightly improved for the N-type region when moving towards the positive voltages.

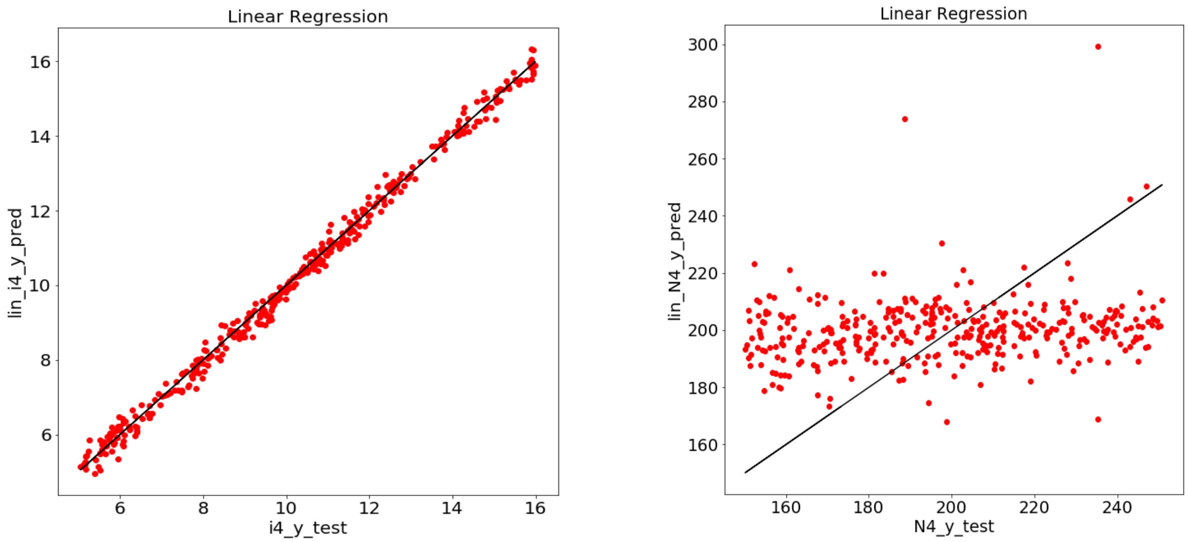

Figure 5: Linear Regression predictions on intrinsic (left) and n-type (right) thicknesses for current values in the region between -2V to 0V.

5.1.3 Training on I (0V to +2V) dataset:

The predictions of n-type and intrinsic layer thicknesses in the forward bias region can be seen in figure 6. The Linear Regression is performing pretty good in the forward bias region.

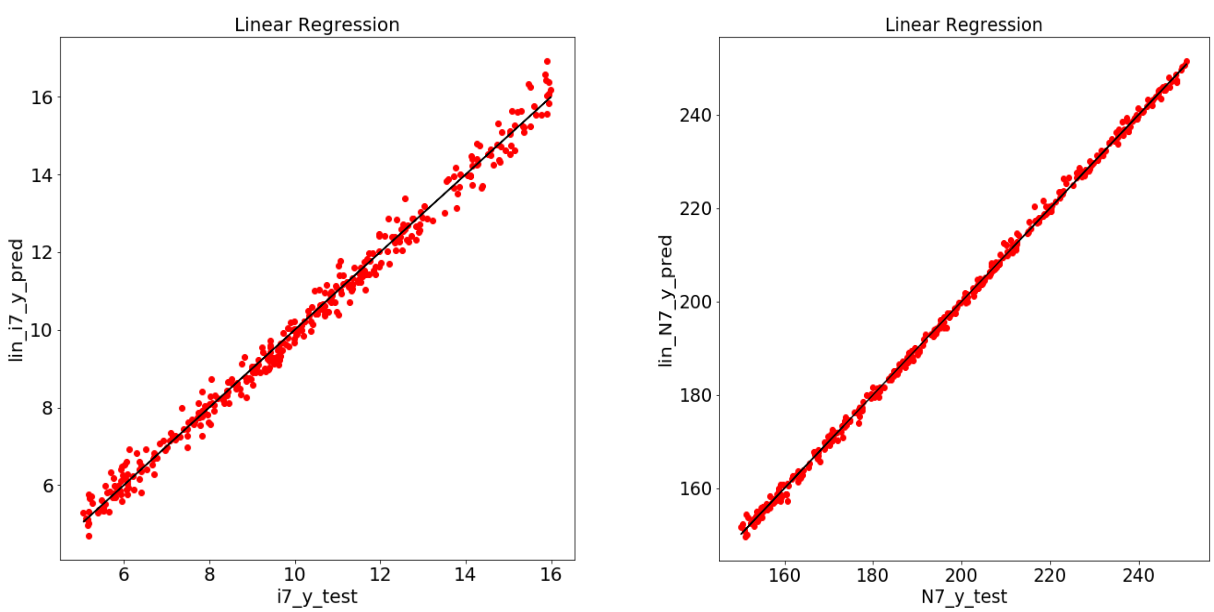

Figure 6: Linear Regression predictions on intrinsic (left) and n-type (right) thicknesses for current values in the region between 0V to +2V.

5.1.4 Training on I (-2V to +2V) dataset:

The entire dataset (current values from -2V to +2V) is used for the training the Linear Regression model. When the wider voltage range is used the accuracy of the predictions is improved. Figure 7 shows the predictions of n_thickness and i_thickness when the model is trained with current values in the region -2V to +2V.

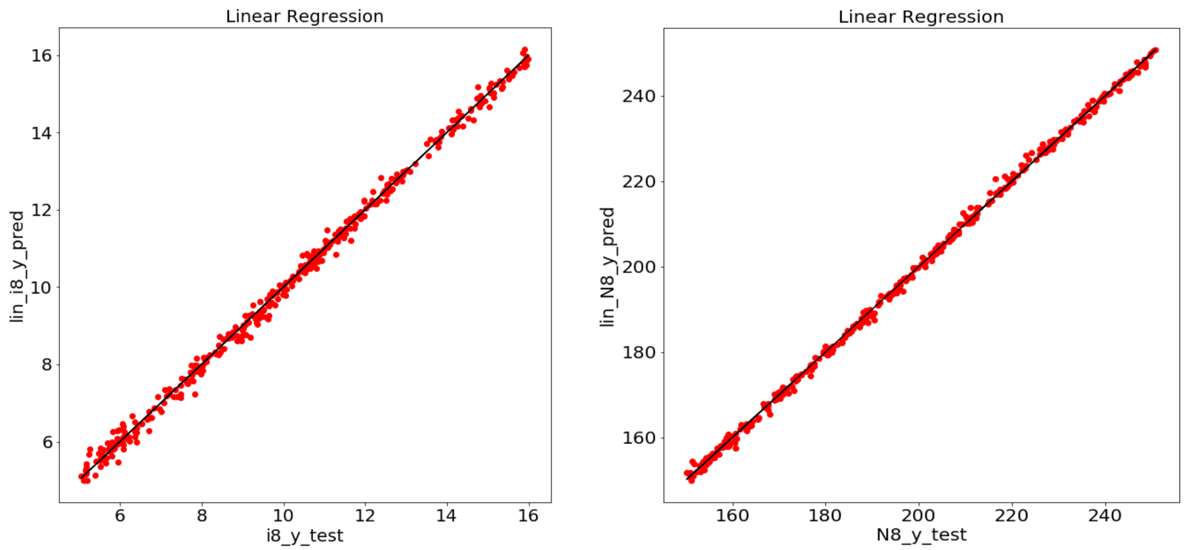

Figure 7: Linear Regression predictions on intrinsic (left) and n-type (right) thicknesses for current values in the region between -2V to +2V.

5.2 Predictions for N-type region in reverse bias and forward bias:

Figure 8 shows the comparison between the Linear Regression predictions of n_thickness for the reverse bias region and forward bias region. This comparison tells about the significance to perform the simulations in regime where the relevant physics is captured.

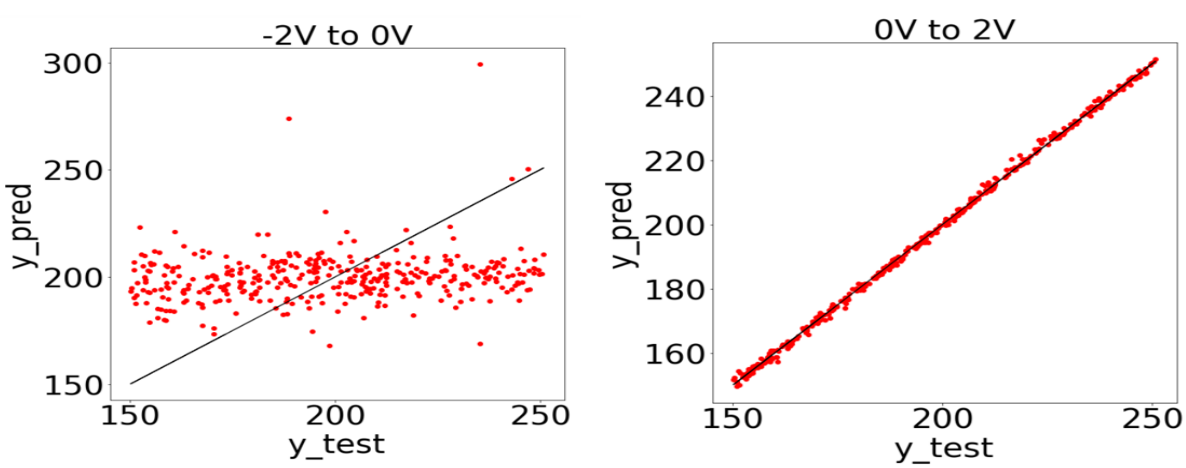

Figure 8: Linear Regression predictions of n_thickness for the reverse bias region (left) and forward bias region (right)

6. 3D TCAD SIMULATIONS USING MACHINE LEARNING:

So far, in this project we have demonstrated 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. For 1D TCAD simulations, the total time taken for each simulation (including the process simulation and device simulation) is 90 seconds approximately. For the complete data generation, it takes around 48 hours for 1D TCAD simulations. For 3D TCAD simulations including process simulation and device simulation, the total time taken to complete one simulation is around 1 hour to 6 hours. If the same study is applied in the industry with thousands of cores, we can expect that the complete data generation can be completed in 24 hours approximately. The simulation time can be further reduced by decreasing the number of datapoints or by decreasing the range of defect variation. The following figure 9 gives an example which shows the predictions with a lesser number of datapoints. From figure 9 we can observe that both Linear Regression and MLP algorithms are performing well with lesser number of datapoints.

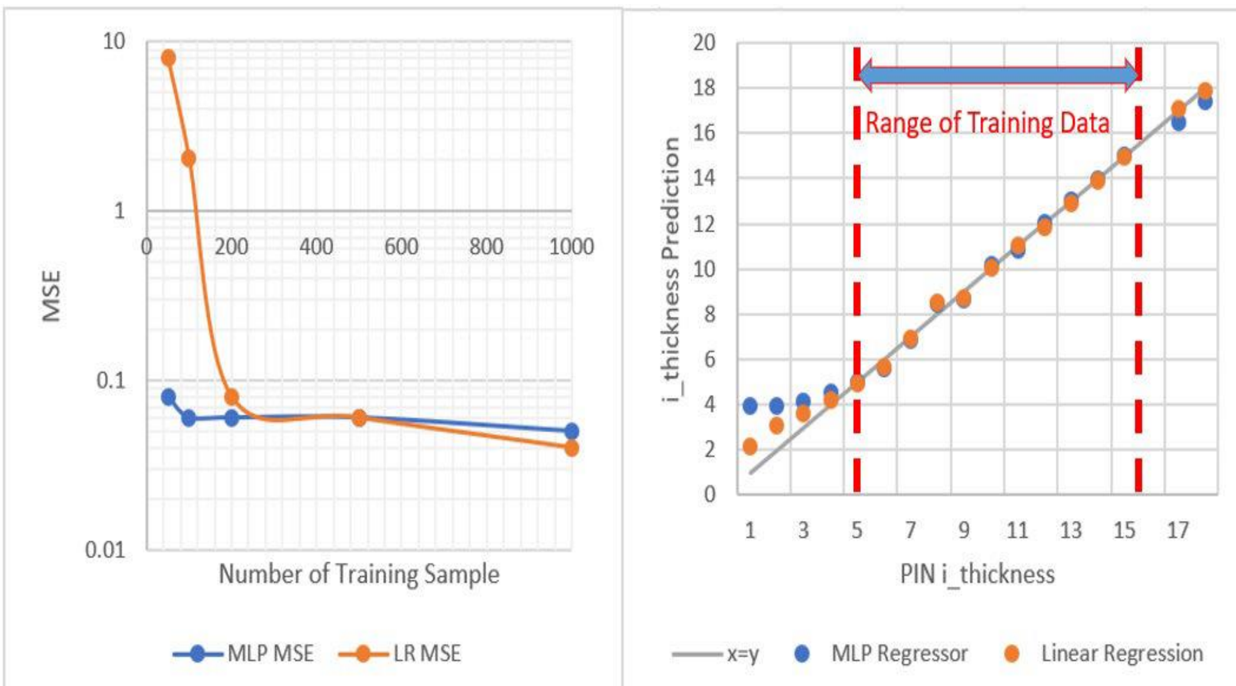

Figure 9: MLP and Linear Regression MSE as a function of number of datapoints (left). Predicted thickness values of intrinsic region as a function of PIN i_thickness (right).

7. SUMMARY AND CONCLUSIONS:

By varying the N-type layer and intrinsic layer thicknesses of a PIN diode, a large amount of data can be generated using TCAD. This data can be used to train the machine learning model to identify the defects based on IV curves. The predictions for N-type layer thickness is good for the positive bias. However, the model performance for N-type layer thickness in the negative bias is poor. In a semiconductor manufacturing industry with thousands of cores, this study can be applied on 3D structures. Moreover, this study can be implemented with lesser number of datapoints for 3D simulations.

8. REFERENCES:

[1] https://pdfs.semanticscholar.org/021e/c6feab2cf17eff0261f1193da92fa3224008.pdf

[2] https://aip.scitation.org/doi/pdf/10.1063/1.4999875

[3] 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

[4] https://www.globalspec.com/learnmore/industrial_engineering_software/engineering_scientific_software/technology_computer_aided_design_software_tcad

[5] https://www.synopsys.com/silicon/tcad/process-simulation.html

[6] https://www.synopsys.com/silicon/tcad/device-simulation.html