Machine Learning Assisted Statistical Variation Analysis of Ferroelectric Transistors: From Experimental Metrology to Predictive Modeling

Gihun Choe[1], Prasanna Venkatesan Ravindran[1], Anni Lu[1], Jae Hur[1], Maximilian Lederer[2], André Reck[2], Sarah Lombardo[3], Nashrah Afroze[1], Josh Kacher[3], Asif Islam Khan[1,3] and Shimeng Yu[1]
[1]School of ECE, Georgia Institute of Technology, Atlanta, GA, USA. [2]Fraunhofer IPMS, Dresden, Germany. [3]School of MSE, Georgia Institute of Technology, Atlanta, GA, USA. Emails: asif.khan@ece.gatech.edu, shimeng.yu@ece.gatech.edu

2022 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits) | 978-1-6654-9772-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/VLSITechnologyandCir46769.2022.9830392

Abstract

We proposed a novel machine learning (ML)-assisted methodology to analyze the variability of ferroelectric field-effect transistor (FeFET) with raw data from the metrology. Transmission Kikuchi diffraction (TKD) measurement was performed on grown Si-doped HfO_2 (Si:HfO_2) thin film. An experimentally acquired polarization map was employed to generate the polarization variation of a ferroelectric gate stack. FeFETs with the multi-domains are simulated in TCAD to generate the training dataset. We trained a neural network using the polarization maps as inputs and the high/low threshold voltage, on-state current, and subthreshold slope as outputs. The trained model with 3,000 data points shows >98% of accuracy and is more than 10^6 times faster than performing TCAD to obtain statistics for 10,000 test samples.

Keywords: Ferroelectrics, Si:HfO_2, FeFET, variation, TKD metrology, TCAD, machine learning, neural network

Introduction

In advanced semiconductor device design, ML-inspired approaches have been suggested to predict data, speed up the simulation, and analytically explain the results [1]. We propose an ML-based predictive modeling framework of FeFET parameter fluctuations caused by the ferroelectric polarization variation. Based on the experimental polarization data extracted from TKD measurement in the SEM metrology [2], we analyzed a FeFET structure at 28 nm node, and generated training and test datasets using Sentaurus TCAD [3] for a multilayer neural network. As is known, it is time consuming (~days) to run 3D TCAD for tens of thousands of samples, while the ML prediction could predict the statistics of the primary device parameters in seconds. The proposed approach could be used to screen the materials, which correlates the thin film properties to the eventual device-level figure of merits.

Methods

A. Experimental Polarization Map by TKD Measurement

The intragranular misorientation angle against the out-of-plane direction of the 10 nm-thickness Si:HfO_2 is gauged by TKD to evaluate the impact of polarization variation on the FeFET (see **Fig. 1**). Assuming the ferro-elastic switching (90° domain wall movement), the b- and c-axis of the domains are switched to simulate a fully woken-up state after measuring a pristine sample. The amplitude of remnant polarization is assumed as 20 μC/cm^2, and it is projected onto the z-axis (out-of-plane). Since each pixel on the figure corresponds to 4nm×4nm, we generate a FeFET structure at 28 nm node with 7×7 domains of the ferroelectric layer for TCAD simulation.

B. TCAD Simulation for Generating Training Dataset

Fig. 2 shows the simulated FeFET bulk structure used in Sentaurus TCAD. Based on the map from TKD data, 14,700 material parameter files are generated for the polarization variation. Those files are assigned to each sample, thereby a sample has its own unique polarization distribution (see **Fig. 3**) [4]. We employed the Preisach model that is built-in with Sentaurus to simulate the multi-domain ferroelectric switching.

C. ML-assisted Acceleration for Variation Analysis

Fig. 5 shows the simulation flow for the ML-based process. To train the ML model for variability analysis of FeFET, the key device features [high/low threshold voltage (HVT/LVT), subthreshold slope (SS), and on-state current (I_{on})] are extracted from TCAD results as training datasets for the supervised learning. After the data generation and refinement, the 2-D tensor of 7×7 is converted into the 1-D (1×49) tensor to feed the neural network. The batch size for input is 50. Next, fully connected neural networks which have 6 hidden layers, 49 inputs, and 4 outputs are designed (see **Fig. 6**). We used leaky ReLU as an activation function of hidden layers. Lastly, stochastic gradient descent is used as the optimizer for minimizing the objective function.

Results

Fig. 7 shows the comparison between HVT and LVT predicted by ML and from TCAD when the number of training datasets is 3,000. To quantitatively assess how well the trained model could predict device features, TCAD results vs. ML-based results which have 500 test samples are compared in **Fig. 8**. The black line denotes the ideal target where ML prediction and TCAD are equal. As the number of training datasets increases, the slope of the fitted trend line is closer to the target line. The coefficient of determination (= R-squared, defined as $1-\Sigma(\hat{y}_i-\bar{y})^2/\Sigma(y_i-\bar{y})^2$, where y_i, and $\hat{y}_i$ are predicted and TCAD value, respectively. $\bar{y}$ is average of y.) also approaches the ideal value 1.0 (see **Fig. 9**). With 3,000 training data points, the model prediction achieves satisfactory weighted mean absolute percentage error (WMAPE = $\Sigma|\hat{y}_i-y_i|/\Sigma|y_i|$, less than 2.1% for HVT, LVT, Ion and SS). The notable advantage of ML-based analysis is the time consumption. Once the neural network is trained (within 0.53 hours for 3,000 data points using CPU), it only takes 2.82 secs to predict the key performances of 14,700 samples. The estimated statistics are reliably close to the TCAD results (see **Table II**). If running TCAD for all these 14,700 samples, it may take 1,146 hours (= 48 days).

Conclusion

ML-assisted FeFET variation study using experimental polarization data has been demonstrated in this work. The neural network model could estimate each device feature with 1.7% of the average error percentage. The time consumption exploiting ML prediction is more than 10^6 times faster compared to TCAD analysis. The proposed methodology could be in principle applied to technology pathfinding with other types of metrology data.

Acknowledgments This work is supported in part by TSMC.

References

[1] C. Jeong *et al*., *IEEE TED*, 68, 11, pp. 5364, 2021. [2] M. Lederer *et al*., *APL*, 115, 222902, 2019. [3] *Sentaurus Device user Guide*, Synopsys, 2017. [4] G. Choe *et al*., *IEEE JEDS*, 9, pp.1131, 2021

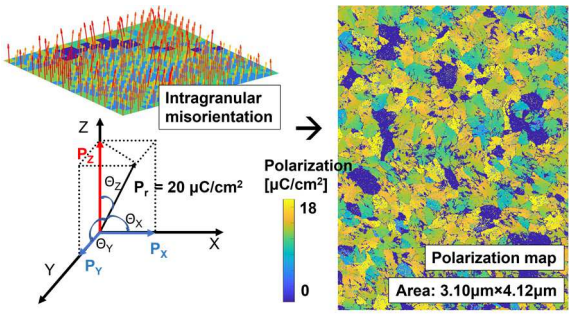


Fig. 1 Out-of-plane polarization extraction method from the phase angle (intragranular misorientation) of the TKD measurement for the polarization map of Si:HfO_2 (Hf:Si= 16:1). The ferroelectric layer is annealed at 1000°C.

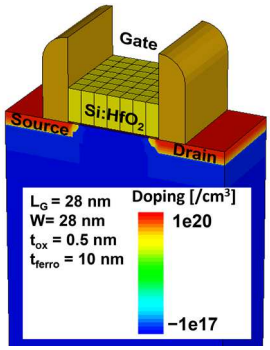


Fig. 2 FeFET structure for TCAD simulation at 28nm.

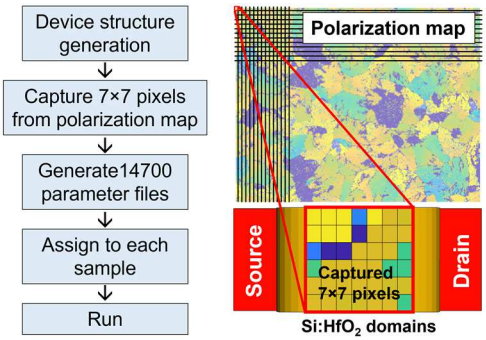


Fig. 3 Flowchart for applying the polarization variation to generate samples for training datasets.

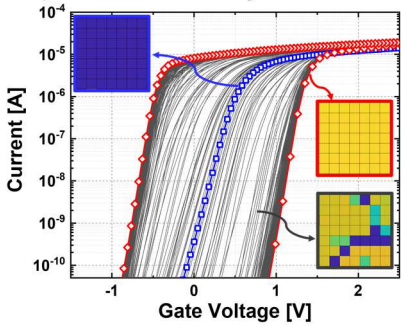


Fig. 4 I_D-V_G of 300 samples out of 14,700 FeFETs. Inside maps are the polarization map corresponding to the indicated curves. Red and blue curves have only ferroelectric phase, and only dielectric phase, respectively.

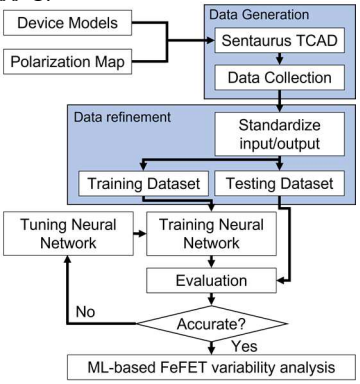


Fig. 5 Machine learning simulation flow for FeFET variability analysis induced by the polarization variation.

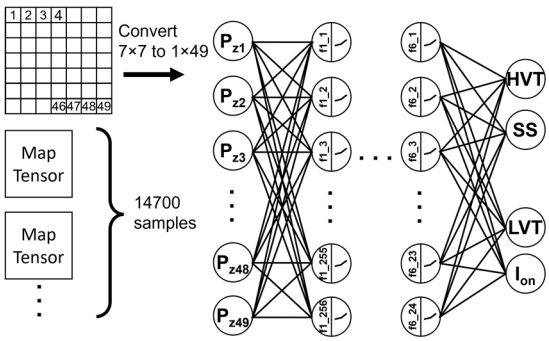
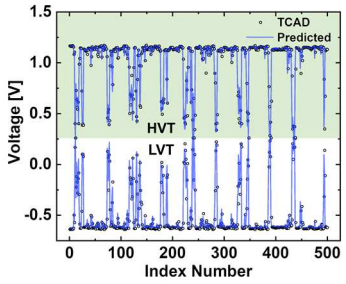


Fig. 6 Fully-connected neural network structure (49-256-128-96-64-48-24-4) used for the ML-based prediction from phase map to device features. Leaky ReLU is used as an activation function.



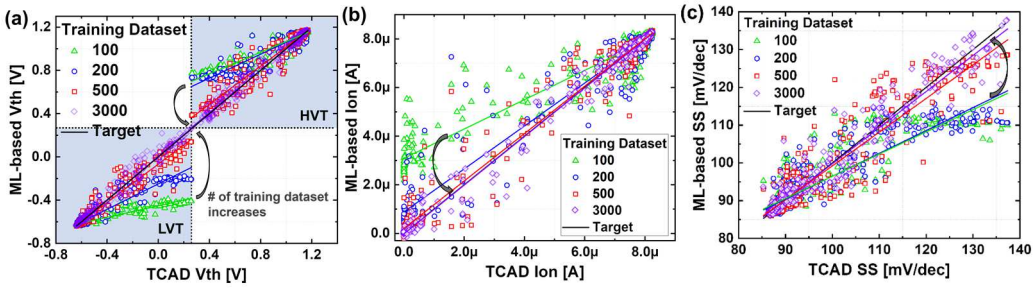


Fig. 7 Predicted HVT/LVT and TCAD-simulated HVT/LVT results of 500 samples. Dots are TCAD test data, lines are ML prediction.

Fig. 8 TCAD results vs. ML-based results [(a) HVT and LVT, (b) I_{on}, (c) SS] which have 500 test samples for each device feature regression. The number of training datasets is varied from 100 to 3,000. Symbols represent the data points and lines are linearly fitted trend lines. For all the features, the fitting curves become closer to the target as the number of training datasets increases. 3,000 training data points are sufficient.

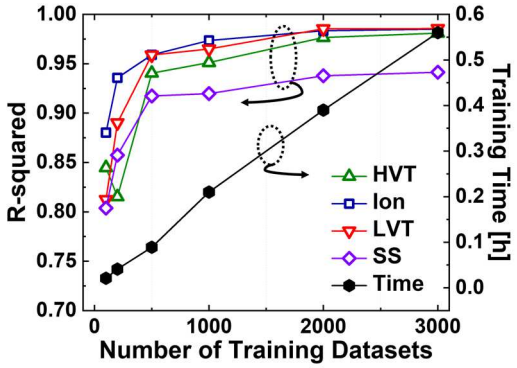


Fig. 9 Left y-axis is the R-squared [coefficient of determination, ideally to be 1, defined as $R^2=1-\Sigma(\hat{y}_i-\bar{y})^2/\Sigma(y_i-\bar{y})^2$, where y_i, and $\hat{y}_i$ are predicted and TCAD value, respectively. $\bar{y}$ is average of y.] and right y-axis is the training time depending on the number of datasets for training, while the number of test samples is 10,000. Training is done using a CPU.

TABLE I. Time consumption after running test samples

# of samples	TCAD simulation time [h]	ML-based inference time [h]
1	0.078	5e-8
100	7.8	5e-6
500	39	2.5e-5
1,000	78	5.3e-5
2,000	156	1.07e-4
10,000	780	5.3e-4
14,700	1146.6 (=48 days)	7.79e-4 (=2.8 secs)

TABLE II. Average, 3-sigma and percentage of prediction error

	TCAD	ML-based prediction		
	μ	μ	Error 3σ	WMAPE
HVT [V]	1.05	1.05	0.08	1.2%
LVT [V]	−0.53	−0.53	0.07	2.0%
I_{on} [μA]	6.68	6.68	1.01	2.1%
SS [mV/dec]	94.52	94.48	9.0	1.6%