

기계학습을 활용한 모바일 반도체 제조 공정에서 동작 전압 예측

백인환*·장승우**·김광수***†

*성균관대학교 반도체디스플레이공학과, **성균관대학교 인공지능학과,
***성균관대학교 소프트웨어융합대학

Operating Voltage Prediction in Mobile Semiconductor Manufacturing Process Using Machine Learning

Inhwan Baek*, Seungwoo Jang** and Kwangsu Kim***†

†Department of Semiconductor and Display Engineering, Sungkyunkwan University,
**Department of Artificial Intelligence, Sungkyunkwan University,
***College of Computing and Informatics, Sungkyunkwan University

ABSTRACT

Semiconductor engineers have long sought to enhance the energy efficiency of mobile semiconductors by reducing their voltage. During the final stages of the semiconductor manufacturing process, the screening and evaluation of voltage is crucial. However, determining the optimal test start voltage presents a significant challenge as it can increase testing time. In the semiconductor manufacturing process, a wealth of test element group information is collected. If this information can be controlled to predict the test voltage, it could lead to a reduction in testing time and increase the probability of identifying the optimal voltage. To achieve this, this paper is exploring machine learning techniques, such as linear regression and ensemble models, that can leverage large amounts of information for voltage prediction. The outcomes of these machine learning methods not only demonstrate high consistency but can also be used for feature engineering to enhance accuracy in future processes.

Key Words : Semiconductor manufacturing, Process data, On-product monitoring, Machine Learning, Ensemble model

1. Introduction

에너지 효율이 중요한 모바일 반도체에선 저전력을 위해 동작전압이 다른 여러 회로의 조합으로 구성된다. 따라서 반도체 제조 공정의 마지막 테스트 단계로 동작 전압에 대한 평가가 있다. 가장 최적의 전압을 찾기 위해선 여러 번 동작이 필요하고 그만큼 테스트 시간이 늘어난다. 반도체 제조 공정에서 테스트 시간을 줄이는 건 이익의 극대화를 위해서 가장 우선시되어야 하는 항목이다.

따라서, 최종적으로 사용하게 될 전압과 가장 가까운 테스트 시작 지점을 찾는 게 중요하다.

최적의 동작 전압을 예측하기 위한 데이터로 제조 공정에서 측정되는 반도체 특성에 대한 데이터들이 있다. 그 중에서 반도체 로직 회로의 설계와 공정 변동성 평가를 위해 사용되는 TEG (Test Element Group)를 활용하고자 한다.[1][2][3] 설계 문제를 해결하기 위해 도입된 TEG는 단층, 단품 검사를 위해 사용될 뿐만 아니라, 수율 평가 및 전압 최적화에 활용할 수 있다.

제조 공정은 여러 단계를 통해 만들어지는 다양한 TEG 값들의 사용전압에 비례하는 관계를 가진다. 그리고 TEG

†E-mail: kim.kwangsu@skku.edu

을 많이 확보할수록 정확한 전압 예측이 가능하다. 하지만 어떤 특징들을 TEG로 넣어 측정하고 어떻게 활용할지에 대한 알고리즘은 엔지니어의 많은 노력을 요구한다.

이와 같은 문제를 해결하기 위해 본 논문에서는 머신러닝을 기반으로 TEG를 활용한 전압을 예측을 수행했다. 모든 feature를 활용하는 linear regression 모델부터 특정 feature에 가중치를 부여하는 앙상블 모델을 통해 전압 예측의 정확도를 높였다.

2. Approach

2.1 Dataset Background

동작 전압을 예측하기 위해 여러 종류의 성능을 표현해 줄 수 있는 정보가 필요하다. 특히, TEG 종류로 Static 회로의 Leakage Current 측정값과 Dynamic 회로의 Ring Oscillator 측정값을 예측 모델의 feature로 활용했다.[5]

Leakage Current가 전압과 관련이 있는 이유는 회로의 더 빠른 동작을 위해 반도체 도핑 농도를 높일수록 낮은 Threshold Voltage에서도 동작할 수 있게 되지만, 높은 도핑 농도로 인한 Leakage Current는 커진다.[4] 반대로 저전력을 위해 높은 Threshold Voltage를 사용하면 낮은 전자 도핑 농도로 인해 작은 Leakage Current를 가진다. [5][10] 이와 같이 동작 전압과 연관이 높은 특징을 보이고 있어 예측 모델에 활용하고자 한다.

Ring Oscillator는 일정 수의 인버터를 연속적으로 연결하여 원형으로 구성된 회로로, 특정 클럭 신호 주기를 측정하는 회로다. [2][11] 반도체 공정에 따라 다양한 요소들의 영향을 받는데 단위 회로마다 Ring Oscillator를 측정하여 설계에 맞게 제작되었는지 검증하는 목적뿐만 아니라 성능을 표현하는 요소로써 동작 전압에 반비례하는 관계를 가지기 때문에 전압을 예측하는 feature로써 활용한다.

본 논문에서는 모바일용 반도체 양산 과정에서 측정된 10,000개의 IDS, Ring Oscillator feature 값을 가지는 데이터로 동작 전압을 예측한다.

2.2 Linear Regression

2.2.1 데이터 상관관계 분석

Linear Regression은 대표적인 회귀 분석 기법으로 공정에서 측정된 값들을 다중 변수로 이용하여 전압과의 상관관계를 분석할 수 있다.

본 논문에서는 측정된 Leakage Current 값과 13 종류의 Ring Oscillator를 활용하여 14개 feature로 회로의 동작 전압을 예측했다. 아래의 Fig. 3에 보이는 것처럼 일부 feature(#1, 2, 3, 4, 5, 7, 9) 에선 높은 상관관계를 보이는 경우도 있지만 작은 표준편차를 보이거나 낮은 상관관계를 보이는 feature

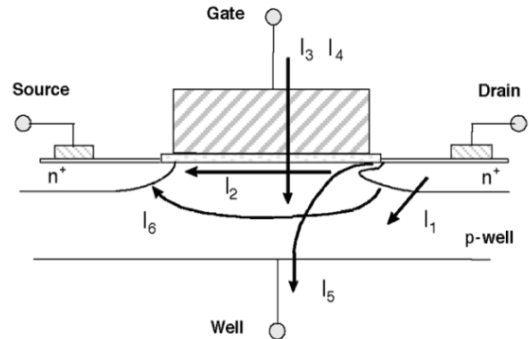


Fig. 1. leakage current in Transistor.

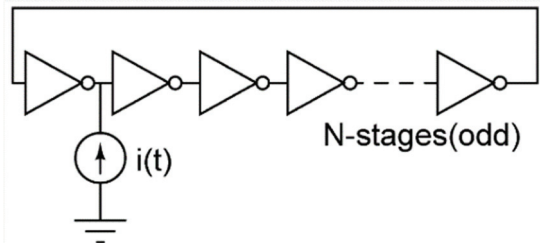


Fig. 2. Current pulse injection in a ring oscillator with arbitrary number of N stages.

(#6, 8, 10, 11, 12, 13, 14)도 있다.

가장 높은 상관관계를 가지는 feature#2 (R2=0.748)를 베이스로 Linear Regression 모델 학습을 진행했다.

2.2.2 Polynomial Features

반도체 전류는 수식을 통해 전압의 제곱에 비례하는 관계임을 알 수 있다.[4]

$$i_D = \mu_n C_{ox} \frac{W}{L} \left[(v_{GS} - V_t) v_{DS} - \frac{1}{2} v_{DS}^2 \right] \quad (1)$$

Threshold Voltage를 기준으로 낮은 전압에서 전류가 지수 함수로 줄어들지만 Threshold Voltage보다 높은 전압으로 가게 되면 거의 변화하지 않으면서 변곡점이 발생한다. 이를 표현하기 위해 Linear regression의 Polynomial Features를 통해 다항식 모델을 만들었다. 또한, 회로는 여러 개의 논리 회로 조합을 통해 이뤄지는 걸 고려하여 각 단위 소자 별 Ring Oscillator feature들의 조합을 반영하기 위해서 intersection 기능이 들어있는 Polynomial 함수를 사용했다.

2.3 Ensemble model

Ensemble model은 단일 회귀만으로 예측 성능이 부족할 때 여러 개의 모델을 조합하여 예측을 하는 기법이다.[13][14]

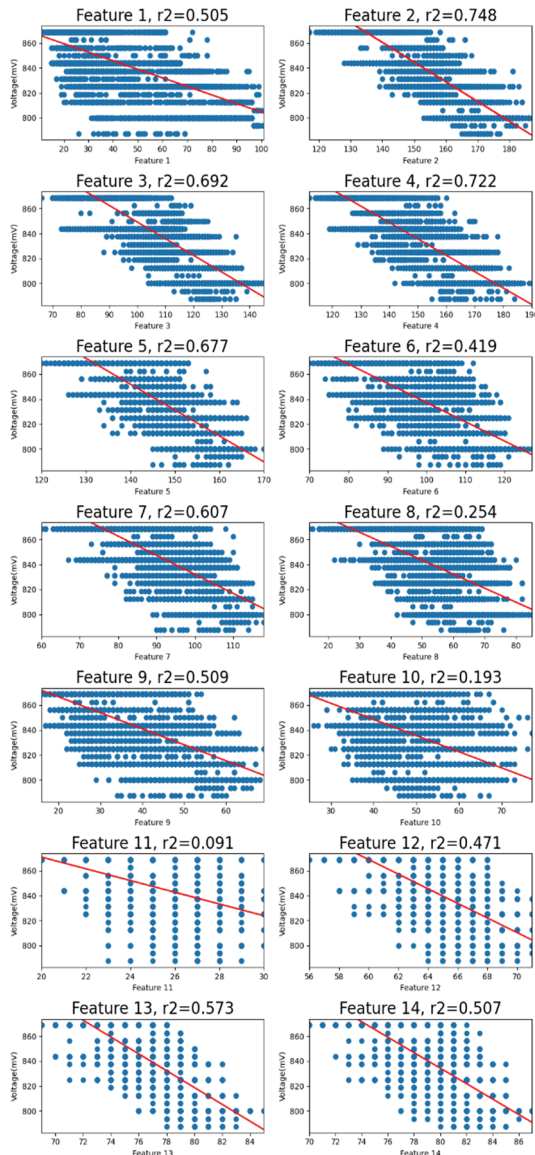


Fig. 3. R2 correlation with target by feature. feature#1 is IDS, feature#2-#16 are different kinds of Ring Oscillator values.

크게 Bagging과 Boosting 방법으로 나뉘는데 Bagging은 다수의 분류기를 학습하여 예측 값들의 평균이나 다수결 투표로 최종 예측을 수행하는 방법이다. 본 논문에서는 대표적으로 많이 사용되는 RandomForest와 ExtraTree 알고리즘을 통해 학습을 진행했다. Random Forest는 의사 결정 나무를 기반으로 한 Ensemble 기법으로, 다양한 특징을 가진 데이터 예측 모델링에 효과적이다.

그리고 Boosting 기법은 이전 학습의 잘못 예측된 샘플을 다음 학습에서 더 잘 예측하도록 수행하는 모델로 본 논문에서는 XGBoost, CatBoost, LightGBM 알고리즘을 사용했다.

이러한 다양한 Ensemble 기법을 활용하여 상호 연관된 방식의 feature들 사이에 어떤 관계를 보이는지 데이터 분석을 통해 일반화 오류를 줄여 보다 정확한 결과를 도출했다.[15]

3. Experiments

3.1 Linear Regression

3.1.1 Implementation Details

최적의 학습을 위해 Cost Function으로 R2, MSE(Mean Squared Error)로 성능 비교를 진행했다. MSE를 사용한 이유는 예측해야 하는 Target 값은 동작 전압으로 실제 값에서 멀어질수록 공정 테스트 시간이 늘어나게 되는데 병렬적으로 동작하는 다른 샘플들의 대기 시간을 감안하여 패널티를 부여할 필요성이 있기 때문이다. 그리고 예측하고자 하는 전압과 같은 단위로 표현하기 위해 RMSE(Root Mean Squared Error)를 사용했다.

그리고 학습을 진행하면서 Train data에 overfitting되는 문제를 예방하기 위해 교차 검증 (Cross validation)으로 Stratified k-fold를 사용했다. 훈련 데이터를 나누면서 일부 데이터 fold에만 특정 전압 샘플이 몰려서 잘못 학습되는 걸 방지하기 위해 Target data의 분포를 고려하여 5-fold로 학습하고 평균값으로 결과들을 비교했다.

3.1.2 Results

다중 feature를 사용한 Linear regression 은 $R^2=0.819$ 으로, 단일 feature를 사용한 Linear regression의 $R^2=0.748$ 보다 높은 결과를 보인다. 하지만 Fig. 4에서처럼 4, 5차 다항식으로 갈수록 과적합(overfitting)되면서 테스트 데이터에서는 성능이 낮아진다. 다항식을 만들 때 차수가 높아질수록 모델의 복잡도가 증가하여 과적합을 막기 위해 intersection only 옵션에 따른 평가에서도 4차 이상에선 동일하게 나타났다. 그래서 Lasso, Ridge 기법을 사용해서 최적화를 진행했다.

Lasso일 때는 $R^2=0.794$, Ridge일 때는 $R^2=0.820$ 으로 정규화를 적용한 모델은 기존 모델과 비슷하거나 약간 떨어지는 결과를 보인다. Linear regression과 동일하게 polynomial 차수는 1차부터 5차까지 실험을 진행하였으며, 학습의 정규화를 조절하는 하이퍼파라미터 alpha 값으로는 0.0001부터 10000까지 로그 스케일로 살펴본 결과 alpha=1에서 최적 값을 보였다.

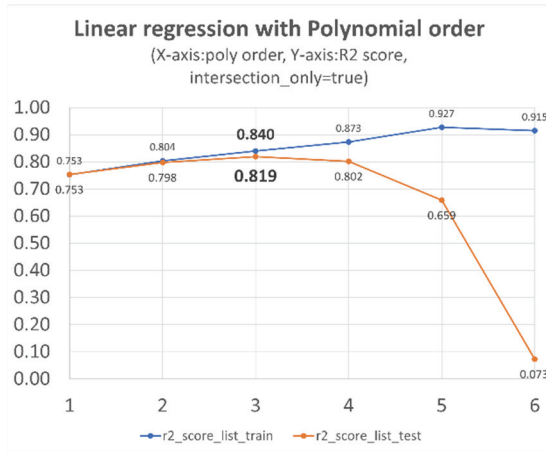


Fig. 4(a). Polynomial experiment results. The best results are obtained with cubic polynomials and intersection only option is True.

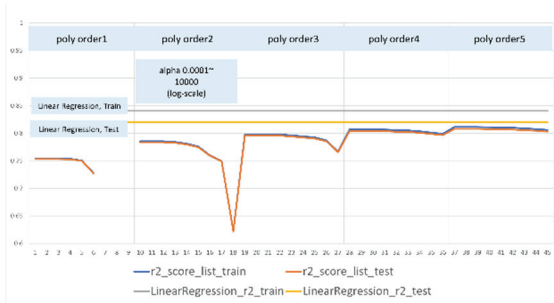


Fig. 5(a). Experiment with Lasso for Regularization.

3.2 Ensemble model

실험을 진행한 12개의 Ensemble model 은 Table 1과 같이 기존 Linear Regression보다 좋은 성능을 보여준다. R2는 Linear regression에 대한 지표이기 때문에 성능 지표로 RMSE, MAE를 사용하여 실제 예측하는 전압과 같은 단위로 비교한 결과에서도 기존 Linear Regression보다 좋은 결과를 보인다.

또한, Table 2는 앙상블 모델을 통해 학습한 모델에서 최종 출력에 각 feature들이 미치는 중요도이다. Weighted Ensemble L2기준으로 feature #2, #1, #3 순으로 중요하게 나왔는데 여기서 feature #1의 경우 Fig.3에서 실제 전압 값과의 관계가 R2=0.505로 낮게 나오지만 전압 예측 모델에선 중요한 것을 확인했다.

Feature importance를 활용하여 예측에 어떤 feature가 가장 영향을 많이 미치는지 그리고 어떤 feature가 필요 없는 것인지 확인했다. 이는 차기 제품에서 정확도를 높이기 위해

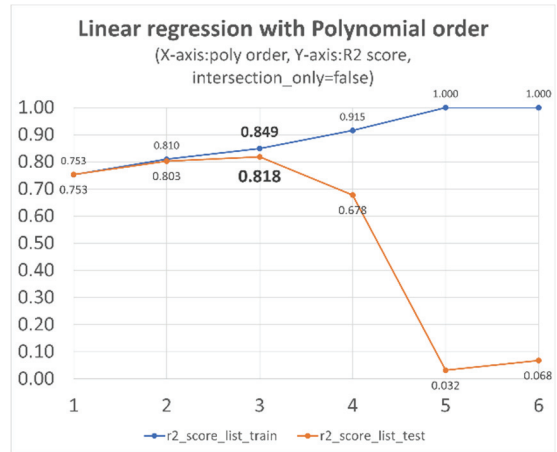


Fig. 4(b). Polynomial experiment results. the intersection only option is False.

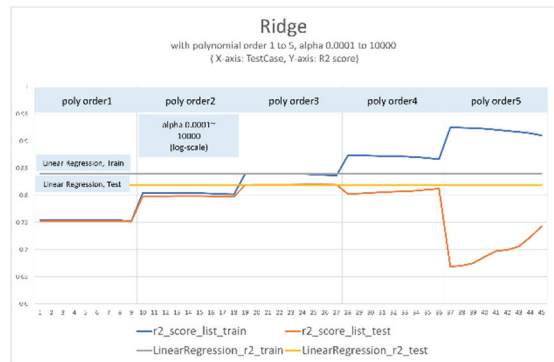


Fig. 5(b). Experiment with Ridge for Regularization.

선 어떤 종류의 feature를 개선하여 차기 공정에서 측정할 지 고민할 중요한 feature engineering으로써 의미를 보여준다.

Table 1. Quantitative results comparison

Model	R2(↑)	RMSE(↓)	MAE(↓)
Linear Regression	0.753	9.859	7.758
Linear Regression_order3th	0.819	8.774	6.807
Lasso	0.794	9.293	7.233
Ridge	0.820	8.887	6.811
WeightedEnsemble_L2	0.918	6.384	3.695
LightGBMLarge	0.916	6.460	3.706
RandomForestMSE	0.916	6.463	3.758
ExtraTreesMSE	0.914	6.530	4.195
XGBoost	0.910	6.681	4.119
CatBoost	0.910	6.691	4.283
LightGBM	0.910	6.708	4.225
LightGBMXT	0.899	7.105	4.807
KNeighborsDist	0.885	7.568	4.291
NeuralNetTorch	0.881	7.704	3.728
NeuralNetFastAI	0.876	7.844	5.664
KNeighborsUnif	0.866	8.161	5.129

Table 2. Feature importance in the ensemble model (Weighted Ensemble L2).

feature index	importance	stddev	p_value	n	p99_high	p99_low
2	144.522	11.681	5.08E-06	5	168.573	120.470
1	68.666	4.946	3.21E-06	5	78.850	58.481
3	51.026	2.068	3.23E-07	5	55.284	46.769
4	4.183	0.922	0.000266	5	6.082	2.285
6	3.131	0.621	0.000176	5	4.409	1.852
5	2.854	0.210	3.50E-06	5	3.287	2.422
14	2.218	0.424	0.000152	5	3.091	1.346
9	1.899	0.492	0.000495	5	2.912	0.886
8	1.759	0.563	0.001103	5	2.918	0.600
10	0.640	0.120	0.000142	5	0.887	0.392
13	0.550	0.274	0.005419	5	1.113	-0.013
7	0.416	0.297	0.017546	5	1.028	-0.195
11	0.189	0.235	0.073262	5	0.673	-0.295
12	-0.045	0.424	0.587198	5	0.828	-0.917

4. Summary

반도체 양산을 진행하며 얻어지는 여러 공정 데이터들로 사용 전압을 예측하여 에너지 효율적인 제품을 위한 목적으로 연구를 시작했다. 각각의 feature들 단독으로 전압을 예측하기 어려웠던 문제를 머신 러닝을 통해, 특히 Ensemble model을 이용함으로써 단일 모델보다 정확한 예측을 할 수 있었다. 더욱 중요한 시사점으로는 feature importance 분석을 통해 모델 예측에 영향이 큰 feature와 작은 feature에 대한 분석이다. 영향도가 높은 feature를 통해 비슷한 계열의 측정값을 늘리고, 낮은 feature 들의 문제점을 개선함으로써 차세대 제품에서 더욱 정확도 높은 모델을 위한 발판을 마련할 수 있었다.

참고문헌

- M. Yamamoto, H. Endo and H. Masuda, "Development of a large-scale TEG for evaluation and analysis of yield and variation," in *IEEE Transactions on Semiconductor Manufacturing*, vol. 17, no. 2, pp. 111-122, May 2004
- A. Bassi, A. Vegetti, L. Croce and A. Bogliolo, "Measuring the effects of process variations on circuit performance by means of digitally-controllable ring oscillators", *Proc ICMTS*, pp. 214-217, 2003.
- Y. Shimizu, M. Nakamura, T. Matsuoka and K. Taniguchi, "Test Structure for Precise Statistical Characteristics Measurement of MOSFETs", *Proc. ICMTS*, pp. 49-54, 2002.
- Pierret and Robert F, "Semiconductor Device Fundamentals" in PEARSON, 2012
- K. Roy, S. Mukhopadhyay and H. Mahmoodi-Meimand, "Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits," in *Proceedings of the IEEE*, vol. 91, no. 2, pp. 305-327, Feb. 2003, doi: 10.1109/JPROC.2002.808156.
- V. De and S. Borkar, "Technology and design challenges for low power and high performance," in *Proc. Int. Symp. Low Power Electronics and Design*, 1999, pp. 163-168.
- S. Thompson, P. Packan, and M. Bohr, "Linear versus saturated drive current: Tradeoffs in super steep retrograde well engineering," in *Dig. Tech. Papers Symp. VLSI Technology*, 1996, pp. 154-155.
- S. Venkatesan, J. W. Lutze, C. Lage, and W. J. Taylor, "Device drive current degradation observed with retrograde channel profiles," in *Proc. Int. Electron Devices Meeting*, 1995, pp. 419-422.
- J. Jacobs and D. Antoniadis, "Channel profile engineering for MOSFET's with 100 nm channel lengths," *IEEE Trans. Electron Devices*, vol. 42, pp. 870-875, May 1995.
- S. Shigematsu, S. Mutoh, Y. Matsuya, Y. Tanabe, and J. Yamada, "A 1-V high speed MTCMOS circuit scheme for power-down applications," *IEEE J. Solid-State Circuits*, vol. 32, pp. 861-869, June 1997.
- Prinzie J, De Smedt V, "Single Event Transients in CMOS Ring Oscillators". *Electronics* 2019, 8, 618.
- Jason Flinn, M. Satyanarayanan. "Energy-aware adaptation for mobile applications" *ACM symposium on Operating systems principles (SOSP '99)*. *SIGOPS Oper. Syst. Rev.* 33, 5 (Dec. 1999), 48-63.
- Bauer, E., Kohavi, R. "An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants." *Machine Learning* 36, 105-139 (1999).
- Geurts, P., Ernst, D. & Wehenkel, L. "Extremely randomized trees." *Machine Learning* 63, 3-42 (2006).
- Ali, K.M., Pazzani, M.J. "Error Reduction through Learning Multiple Descriptions." *Machine Learning* 24, 173-202 (1994).
- Song-Yeon Lee and Yong Jeong Huh, "A Study on Prediction Model of Scaffold Appearance Defect Using Machine Learning", *Journal of the Semiconductor & Display Technology*, Vol.19, no.2, pp. 26-30, 2020.

접수일: 2023년 3월 16일, 심사일: 2023년 3월 22일,
게재확정일: 2023년 3월 22일