Data-Efficient Prediction of Minimum Operating Voltage via Inter- and Intra-Wafer Variation Alignment

Yuxuan Yin[†], Rebecca Chen[‡], Chen He[‡], Peng Li[†]
[†]Department of Electrical and Computer Engineering, University of California Santa Barbara, CA, USA
[‡]Automotive Processing, NXP Semiconductors, TX, USA

***Abstract*—Predicting the minimum operating voltage (V_{min}) of chips stands as a crucial technique in enhancing the speed and reliability of manufacturing testing flow. However, existing V_{min} prediction methods often overlook various sources of variations in both training and deployment phases. Notably, overlooking wafer zone-to-zone (intra-wafer) variations and wafer-to-wafer (inter-wafer) variations diminishes the accuracy, data efficiency, and reliability of V_{min} predictors. To address this challenge, we propose Restricted Bias Alignment (RBA), a novel data-efficient V_{min} prediction framework that introduces a variation alignment technique to simultaneously estimate inter- and intra-wafer variations. Furthermore, we propose utilizing class probe data to model inter-wafer variations for the first time.**

***Index Terms*—chip performance prediction, machine learning, process variation, data alignment**

I. Introduction

The measurement of the minimum operating voltage (V_{min}) represents a pivotal testing procedure crucial for assessing chip performance. It enables the identification of substandard products, facilitates power consumption optimization, and serves as an early indicator of potential failures during the device's lifespan. A case study involving 7nm industry chips illustrates that subjecting all chips to uniform energy levels leads to a minimum 16% increase in energy utilization [1].

Current industrial practices rely on die-level features to construct V_{min} prediction models that account for die-to-die variations. These features, gathered from parametric tests or on-chip monitors such as IDDQ tests and ring oscillators [2], [3], serve as inputs to machine learning-based V_{min} predictors [4]–[8]. However, existing methodologies fall short in capturing wafer zone-to-zone (intra-wafer) and wafer-to-wafer (inter-wafer) variations. Fig. 1a and Fig. 1b illustrate the impact of inter- and intra-wafer variations to V_{min} and parametric features in an industrial 16nm chip dataset, respectively. It is evident that both types of process variation significantly alter the distribution of V_{min} and parametric features, ultimately impairing the accuracy of aforementioned V_{min} predictors.

In this paper, we introduce a novel V_{min} prediction framework called restricted bias alignment (RBA), designed to systematically capture inter- and intra-wafer variations, along with die-to-die variations. To address die-level variations, we adopt parametric test features in line with prior research. However, for inter- and intra-wafer variations, we treat them as independent and employ a voltage bias term to model their respective impacts on individual dies. Additionally, we utilize class probe data to model inter-wafer variations. By aligning and modeling process variations, RBA is data-efficient and robust during the training process, and is accurate in deployment for dies from new wafers. Our main contributions are:

- We propose a novel data-efficient algorithm for estimating and aligning V_{min} shifts resulting from inter- and intra-wafer variations.
- We propose to utilize class probe data for inter-wafer V_{min} shift modeling for the first time, and propose to reuse pre-learned intra-wafer V_{min} shift for dies from new wafers in addressing process variations.
- Through empirical analysis, we demonstrate the effectiveness and data efficiency of the proposed V_{min} prediction approach on an industrial dataset.

II. Preliminaries

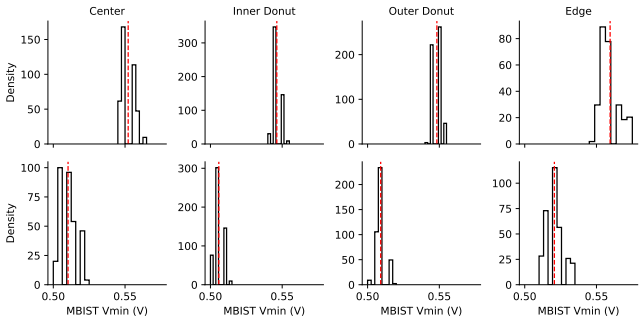
A. Testing Flow in Semiconductor Manufacturing

In semiconductor manufacturing, wafer-level testing employs class probe test structures that are situated in the scribe lines between product dies, outside the actual chips, as depicted in Fig. 2. These test structures typically contain components like transistors, via chains, resistors, and capacitors, mirroring the fabrication process used for the product dies. The purpose of these test structures is to provide feedback on the wafer's processing, enable statistical process control, and help reduce variations from wafer to wafer.

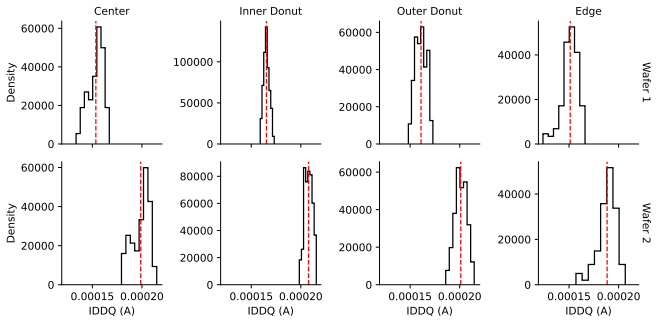
At the die level, each part incorporates its own set of test structures, known as Process Observation Structures (POSt), typically located at the corners of each die. These structures include components like ring oscillators, transistors, bipolar junctions, resistors, and capacitors, which reflect the elements used in the actual circuits on each die. POSt structures offer visibility into the die-level processing, allowing engineers to monitor and evaluate the performance of individual dies.

Together, these two sets of test structures—the class probe at the wafer level and the POSt at the die level—create a hierarchical system for tracking process variations across wafers and individual dies. This setup allows for correlations between class probe data and inter-wafer Vmin bias, as well as between POSt data and the Vmin of individual dies. The presence of these correlations indicates that the test structures can be used

979-8-3315-2144-8/25/$31.00



(a) MBIST V_{min} variation



(b) IDDQ current variation

Figure 1: Wafer-to-wafer and wafer zone variations across 2 wafers, measured at 25°C. Red dashed lines represent mean values.

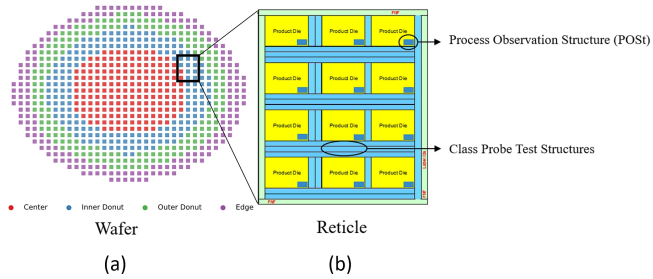


Figure 2: (a) A balanced wafer region partition into 4 regions; (b) Class probe test structures and process observation structures

to predict and control process variations, contributing to more reliable and consistent chip production.

Table I: Inference Time Comparison of V_{min} Regressors

Model	Inference Time (ms)
Linear Regression (LR)	0.167
XGBoost [9]	1.706
CatBoost [10]	20.264

B. Selections of V_{min} Predictor

We firstly discuss the selection of regression models for predicting the V_{min} point, leveraging insights from a recent paper [11]. A variety of regressors were evaluated by the authors, including Linear Regression (LR), ensemble tree-based methods such as XGBoost [9] and CatBoost [10], as well as Neural Networks.

The analysis of the authors of [11], conducted on 5nm industrial chips, revealed no universally optimal model across all scenarios. Importantly, LR's simplicity and significantly faster inference time make it particularly attractive for industrial applications, where inference time directly impacts the overall manufacturing test time, which is subject to stringent constraints. Table I summarizes the inference time for the evaluated models on our 16nm automotive chips. LR achieves inference times that are approximately 10× faster than XGBoost and 100× faster than CatBoost. This significant difference highlights LR as the most efficient option when balancing prediction accuracy and V_{min} test time.

Given these considerations,we focus on improving LR for V_{min} prediction under process variations, even though the concept is generalizable to other, more complex regressors.

C. Linear Regression for V_{min} Prediction

Linear regression is a simple yet effective method to predict V_{min}. It builds upon the following assumption

$$\mathrm{y} = \mathbf{x}\mathbf{w} + \mathrm{b} + \epsilon \tag{1}$$

where $\mathrm{y} \in (0, +\infty)$ is the positive value of V_{min}, $x \in \mathbb{R}^{1\times d}$ is a d-dimensional row vector, which is a subset of features measured by parametric tests, $\mathbf{w} \in \mathbb{R}^{d\times 1}$ is a d-dimensional column vector of unknown parameters, $\mathrm{b} \in \mathbb{R}$ is a bias term of V_{min}, and $\epsilon \in \mathbb{R}$ accounts for the influence on V_{min} other than features $\mathbf{x}$.

Given a training dataset $(\mathbf{X}, \mathbf{y})$, one can estimate $\widehat{\mathbf{w}}$ and $\widehat{\mathrm{b}}$ via minimizing the sum of square residuals

$$\widehat{\mathbf{w}}, \widehat{\mathrm{b}} = \underset{\mathbf{w},\mathrm{b}}{\arg\min} ||\mathbf{y} - \mathbf{X}\mathbf{w} - \mathrm{b}||_2^2 \tag{2}$$

and the solution is

$$\widehat{\mathbf{w}} = \left(\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}^T\widetilde{\mathbf{y}}, \quad \widehat{\mathrm{b}} = \bar{\mathbf{y}} - \bar{\mathbf{X}}\widehat{\mathbf{w}} \tag{3}$$

where the bar operator $\bar{\cdot}$ computes the mean value (vector) of a vector (matrix), and the tilde operator $\widetilde{\cdot}$ centralizes the input.

D. Influence of Inter- and Intra-Wafer Variation on V_{min} Prediction

Process variations are inherent in modern semiconductor manufacturing, with their significance magnifying as technology nodes and wafer sizes scale. Typically, there are two types of process variations: inter-wafer (wafer-to-wafer) variations and intra-wafer (zone-to-zone) variations.

We visually depict the contributions of both variations to the distribution of V_{min} and parametric features within our industrial 16nm automotive dataset in Fig. 1. Specifically, we present histogram plots and mean values of MBIST V_{min} in Fig. 1a and IDDQ current in Fig. 1b, spanning the four regions of two wafers from the same lot. Each row in either sub-figure represents the intra-wafer variation of a given wafer, while each

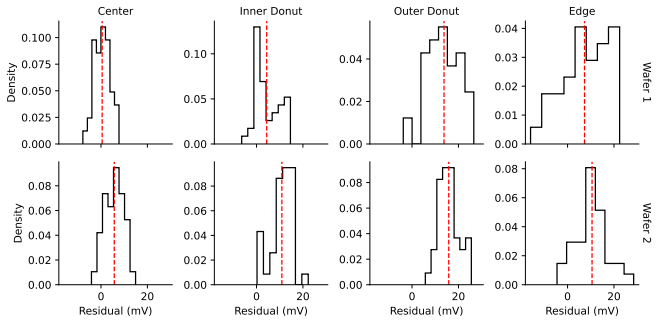


Figure 3: Variation of testing residuals of V_{min} prediction of a linear regression trained on dies from the center zone from wafer 1. Red dashed lines represent mean values.

column signifies the inter-wafer variations of a specific wafer zone.

It is evident that both V_{min} and parametric feature distributions exhibit considerable variance across wafers and regions. However, if process variations merely introduce *covariate shift*, wherein the relationship $\mathrm{y}|\mathbf{x}$ between V_{min} and parametric features remains constant, we could feasibly train a V_{min} predictor and deploy it on new testing dies. Unfortunately, the assumption of covariate shift does not hold for the V_{min} prediction task. To illustrate, we train the aforementioned linear model on dies from the center zone of wafer 1 and test it across all four zones of wafers 1 and 2. The resulting residual of MBIST V_{min} on the testing data is depicted in Fig. 3, where the residual r is computed as $\mathrm{r} = \mathrm{y} - \widehat{\mathrm{y}}$.

The predictor performs relatively well on the center zone of wafer 1; however, its accuracy notably declines on other testing wafer zones of both wafer 1 and wafer 2. This outcome underscores that process variations alter the statistical correlation between V_{min} and parametric test features, rather than solely inducing covariate shift. Consequently, there arises a necessity to systematically address process variations, encompassing both inter- and intra-wafer variations, to attain accurate and robust V_{min} prediction.

III. Data Efficient Inter- and Intra-Wafer Variation Alignment

It is clear that both V_{min} and parametric features exhibit significant variation from wafer to wafer and from region to region. In this paper, we concentrate on a specific impact of process variations on the dependency $\mathrm{y}|\mathbf{x}$ between V_{min} and parameter features: a consistent voltage shift of V_{min} relative to the bias b.

We introduce Restricted Bias Alignment (RBA) to align the V_{min} shift resulting from process variations. RBA operates under the assumption that intra-wafer and inter-wafer variations are independent. This assumption is motivated by observations from Fig. 1, where the intra-wafer variation appears consistent across wafers, and the inter-wafer variation remains stable across wafer regions.

Moreover, RBA introduces the use of class probe features to model inter-wafer variations. By leveraging the pre-learned

Zone	Center	Inner Donut	Outer Donut	Edge
# Dies	180	184	184	183

Table II: The number of dies in each wafer zone

intra-wafer variations and the predicted inter-wafer variations, RBA demonstrates the capability to predict V_{min} for a die from a new wafer without necessitating further data collection.

A. Restricted Bias Alignment (RBA) for V_{min} Prediction

1) Problem Formulation: Denote $\mathbf{z}_i \in \mathbb{R}^k$ as the vector of k class probe features of the i-th wafer, RBA models V_{min} as

$$\mathbf{y}_{i,j} = \mathbf{X}_{i,j}\mathbf{w} + \mathrm{b}_i^{inter}(\mathbf{z}_i) + \mathrm{b}_j^{intra} + \epsilon \tag{4}$$

where b_j^{intra} accounts for the V_{min} shift caused by intra-wafer variations; $\mathrm{b}_i^{inter}(\mathbf{z}_i)$ is a linear model of $\mathbf{z}_i$, representing the voltage bias of wafer i:

$$\mathrm{b}_i^{inter}(\mathbf{z}_i) := \mathbf{z}_i\mathbf{w}_{\mathbf{z}} + \mathrm{b}_{\mathbf{z}} + \epsilon_{\mathbf{z}} \tag{5}$$

We construct a loss function $\mathcal{L}_{RBA}$ as the sum of square residuals of the V_{min} prediction across the whole training set:

$$\mathcal{L}_{RBA} := \sum_{i,j} ||\mathbf{y}_{i,j} - \mathbf{X}_{i,j}\mathbf{w} - \mathrm{b}_i^{inter} - \mathrm{b}_j^{intra}||_2^2 \tag{6}$$

where $\mathbf{b}$ is a set of V_{min} shift of inter- and intra-wafer variations. We minimize the loss function it to estimate $\mathbf{w}$, $\mathbf{b}^{inter}$, and $\mathbf{b}^{intra}$:

$$\widehat{\mathbf{w}}, \widehat{\mathbf{b}}^{inter}, \widehat{\mathbf{b}}^{intra} = \underset{\mathbf{w},\mathbf{b}^{inter},\mathbf{b}^{intra}}{\arg\min} \mathcal{L}_{RBA}\left(\mathbf{w}, \mathbf{b}^{inter}, \mathbf{b}^{intra}\right) \tag{7}$$

2) Solution: To directly solve Eq. (7) is complicated. We adopt an alternative one-step gradient descent approach, integrating a novel initialization method to accelerate the convergence. Denote $\mathbf{w}_{(t)}$, $\mathbf{b}_{(t)}^{inter}$, and $\mathbf{b}_{(t)}^{intra}$ as the estimated parameters in the training step t.

In step t, we first optimize $\mathbf{w}_{(t)}$ to minimize $\mathcal{L}_{RBA}$, condition on $\mathbf{b}_{(t-1)}^{inter}$ and and $\mathbf{b}_{(t-1)}^{intra}$:

$$\mathbf{w}_{(t)} = \underset{\mathbf{w}}{\arg\min}\, \mathcal{L}_{RBA}\left(\mathbf{w}, \mathbf{b}_{(t-1)}^{inter}, \mathbf{b}_{(t-1)}^{intra}\right) \tag{8}$$

This is a linear regression problem and the solution is

$$\mathbf{w}_{(t)} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}_{(t)} \tag{9}$$

where $(\mathbf{X}, \mathbf{y}_{(t)})$ is the concatenation of data of all wafer zones $(\mathbf{X}_{i,j}, \mathbf{y}_{i,j} - \mathrm{b}_{i;(t-1)}^{inter} - \mathrm{b}_{j;(t-1)}^{intra})$.

Then, we adopt the chain rule to update biases:

$$\mathbf{b}_{(t)}^{inter} = \mathbf{b}_{(t-1)}^{inter} - \eta\left(\frac{\partial\mathcal{L}_{RBA}}{\partial\mathbf{w}_{(t)}} \cdot \frac{\partial\mathbf{w}_{(t)}}{\partial\mathbf{b}_{(t-1)}^{inter}}\right)^T \tag{10}$$

$$\mathbf{b}_{(t)}^{intra} = \mathbf{b}_{(t-1)}^{intra} - \eta\left(\frac{\partial\mathcal{L}_{RBA}}{\partial\mathbf{w}_{(t)}} \cdot \frac{\partial\mathbf{w}_{(t)}}{\partial\mathbf{b}_{(t-1)}^{intra}}\right)^T \tag{11}$$

where η is a hyper-parameter of learning rate.

Once the training process is done, we optimize the coefficients of class probe features:

$$\widehat{\mathbf{w}}_{\mathbf{z}}, \widehat{\mathrm{b}}_{\mathbf{z}} = \underset{\mathbf{w}_{\mathbf{z}},\mathrm{b}_{\mathbf{z}}}{\arg\min} \sum_i ||\widehat{\mathrm{b}}_i^{inter} - \mathbf{z}_i\mathbf{w}_{\mathbf{z}} - \mathrm{b}_{\mathbf{z}}||_2^2 \tag{12}$$

Table III: The testing RMSE (mV) of DC Scan V_{min} prediction with 75% data for training

Temperature		-45°C				25°C			
Wafer ID	Wafer Zone	# Die	Linear Model	BA (Reference)	Restricted BA	# Die	Linear Model	BA (Reference)	Restricted BA
1	Center	166	3.54	2.69	2.88	165	3.28	2.72	2.68
1	Inner Donut	162	5.35	3.71	4.14	161	5.03	4.12	4.64
1	Outer Donut	158	7.69	3.24	3.28	173	4.79	3.92	3.91
1	Edge	91	5.21	3.34	3.05	128	5.17	4.82	4.78
2	Center	152	11.84	3.80	4.90	166	6.13	3.63	3.98
2	Inner Donut	153	8.23	3.57	3.85	172	5.16	3.62	3.79
2	Outer Donut	155	5.04	3.25	3.59	172	4.37	4.03	3.89
2	Edge	93	9.39	5.60	5.58	128	7.82	5.51	5.72
3	Center	54	8.03	4.46	4.37	152	5.49	4.36	4.38
3	Inner Donut	111	5.19	3.05	3.49	168	3.98	3.06	3.46
3	Outer Donut	126	5.74	4.52	4.69	173	4.50	4.43	4.72
3	Edge	89	4.03	4.07	4.27	130	4.66	4.68	4.71
4	Center	138	4.25	3.73	3.60	161	3.58	2.43	2.38
4	Inner Donut	157	3.98	2.72	2.72	174	3.63	3.32	3.46
4	Outer Donut	154	5.16	2.92	2.94	171	5.61	3.93	3.88
4	Edge	91	3.52	3.56	3.55	128	5.18	4.58	4.88
5	Center	153	5.98	3.12	3.49	171	3.80	2.70	3.02
5	Inner Donut	158	4.34	3.12	3.11	172	4.23	3.14	3.43
5	Outer Donut	160	6.11	3.40	3.44	178	6.15	5.36	5.28
5	Edge	104	4.40	4.27	4.32	141	4.23	4.25	4.25
Mean		-	6.20	**3.56**	<u>3.75</u>	-	4.91	**3.97**	<u>4.10</u>

Table IV: The testing RMSE (mV) of AC Scan V_{min} prediction with 75% data for training

Temperature		-45°C				25°C			
Wafer ID	Wafer Zone	# Die	Linear Model	BA (Reference)	Restricted BA	# Die	Linear Model	BA (Reference)	Restricted BA
1	Center	166	7.71	5.08	5.01	164	5.49	4.72	4.62
1	Inner Donut	162	6.70	6.33	6.49	159	6.54	6.12	6.20
1	Outer Donut	158	5.29	5.53	5.45	173	8.90	5.79	5.63
1	Edge	90	6.66	5.85	6.19	123	9.43	8.32	8.44
2	Center	152	12.06	5.57	6.00	165	13.69	5.17	6.10
2	Inner Donut	153	7.35	6.77	6.83	170	7.23	6.23	5.81
2	Outer Donut	155	6.10	6.34	6.58	170	6.16	6.29	6.41
2	Edge	93	5.27	4.61	4.48	121	7.66	6.19	5.88
3	Center	54	11.52	6.84	7.22	151	7.94	5.88	6.37
3	Inner Donut	110	8.72	5.82	6.23	165	7.77	6.03	6.25
3	Outer Donut	126	7.82	6.30	6.59	172	6.88	6.78	6.95
3	Edge	89	8.98	8.43	8.84	127	7.24	7.83	8.52
4	Center	138	6.48	4.67	5.10	160	5.78	5.64	5.96
4	Inner Donut	157	9.33	6.59	6.72	172	6.09	4.96	5.48
4	Outer Donut	154	12.21	7.18	7.30	168	11.23	7.29	7.38
4	Edge	91	10.59	6.52	6.65	123	9.07	8.15	8.34
5	Center	153	5.54	5.38	5.01	172	6.53	6.18	6.17
5	Inner Donut	158	6.09	6.31	6.24	170	5.86	5.66	5.78
5	Outer Donut	159	6.95	6.56	6.83	176	9.29	6.22	6.19
5	Edge	103	6.39	6.44	6.89	136	7.30	6.87	6.20
Mean		-	8.05	**6.17**	<u>6.33</u>	-	8.04	**6.31**	<u>6.43</u>

3) Discussion: RBA effectively separates the influence of inter- and intra-wafer variations on V_{min} shift. This decoupling mechanism distinguishes RBA from BA, mitigating potential overfitting concerns particularly when dealing with small training datasets, thereby bolstering overall data efficiency.

For a testing die $(\mathbf{x}_{i,j}^{test}, \mathbf{z}_i^{test}, \mathrm{y}_{i,j}^{test})$ form the j-th zone of the i-th wafer, the V_{min} prediction of RBA is

$$\widehat{\mathrm{y}}_{i,j}^{test} = \mathbf{x}_{i,j}^{test}\widehat{\mathbf{w}} + \widehat{\mathrm{b}}_i^{inter}(\mathbf{z}_i^{test}) + \widehat{\mathrm{b}}_j^{intra} \tag{13}$$

where

$$\widehat{\mathrm{b}}_i^{inter}(\mathbf{z}_i^{test}) = \mathbf{z}_i^{test}\widehat{\mathbf{w}}_{\mathbf{z}} + \widehat{\mathrm{b}}_{\mathbf{z}} \tag{14}$$

By incorporating class probe features to capture inter-wafer variations, RBA possesses the capability for deployment in V_{min} prediction without necessitating re-training or measuring V_{min} for any dies from a new wafer. This feature enhances the practical applicability and efficiency of RBA in product testing scenarios.

IV. Experimental Results

We conduct experiments to demonstrate the efficacy of our approach RBA for addressing inter- and intra-wafer variations on thousands of 16nm automotive chips. We aim to illustrate: 1) the effectiveness of V_{min} bias alignment, 2) the data efficiency and robustness of RBA, 3) the capability of class probe features to capture inter-wafer variation, and 4) the ability of RBA to predict V_{min} of dies from a new wafer.

a) Description of Data Collection: We get the class probe data of each wafer from the foundry. During the testing flow of product manufacturing, V_{min}, including DC Scan V_{min}, AC Scan V_{min}, and MBIST V_{min} are measured at at three different temperatures: -45°C (cold), 25°C (room), and 125°C (hot).

Table V: The testing RMSE (mV) of MBST V_{min} prediction with 75% data for training

Temperature		-45°C				25°C			
Wafer ID	Wafer Zone	# Die	Linear Model	BA (Reference)	Restricted BA	# Die	Linear Model	BA (Reference)	Restricted BA
1	Center	166	8.73	2.58	2.64	165	4.54	2.85	3.40
1	Inner Donut	162	9.91	3.25	3.37	161	8.97	3.53	4.18
1	Outer Donut	158	9.34	4.76	4.72	173	10.10	4.49	4.45
1	Edge	90	13.10	9.76	11.94	120	15.98	9.77	11.17
2	Center	135	6.46	3.69	4.04	166	11.05	3.43	3.49
2	Inner Donut	136	5.19	3.70	3.69	172	8.07	2.68	2.71
2	Outer Donut	80	5.48	3.92	3.50	172	5.62	3.74	3.71
2	Edge	59	10.54	4.83	4.85	124	13.17	7.29	7.08
3	Center	54	8.99	3.35	3.52	152	8.24	4.20	4.10
3	Inner Donut	111	5.54	3.27	3.29	165	5.54	2.66	2.62
3	Outer Donut	125	5.67	3.04	3.05	174	6.70	2.56	2.56
3	Edge	88	9.05	6.69	6.69	123	11.83	7.32	7.34
4	Center	138	30.50	4.23	6.24	161	24.39	4.22	5.68
4	Inner Donut	157	24.24	3.20	3.13	174	14.82	3.08	3.21
4	Outer Donut	153	22.69	3.46	4.35	171	10.59	3.01	3.04
4	Edge	91	28.54	5.12	6.96	128	13.90	5.28	6.32
5	Center	153	19.38	4.62	4.65	172	18.45	4.64	4.89
5	Inner Donut	158	18.20	4.44	4.51	172	24.10	4.31	4.34
5	Outer Donut	159	19.46	4.30	4.42	178	25.02	4.24	4.42
5	Edge	104	13.10	5.80	5.78	138	19.39	6.69	6.77
Mean		-	16.25	**4.45**	<u>4.91</u>	-	14.47	**4.63**	<u>4.94</u>

Similarly, parametric test and POSt test data were collected under different temperatures from Automatic Test Equipment (ATE) testers.

Our dataset has several wafers. Each wafer is partitioned into 4 zones: center, inner donut, outer donut, and edge. The visualization of this partition is shown in Fig. 2, and the number of dies in each wafer zone is listed in Table II. Due to the expensive cost of V_{min} test, only a subset of dies is performed the V_{min} test for a certain test pattern.

b) RBA Settings: RBA leverages 5 parametric test features and 2 class probe features as input to predict V_{min}. All features are selected by the Correlation Feature Selection algorithm [12], and pass the causation check by our testing engineer. The hyper-parameter learning rate η is set to 0.1. We terminate the training process when the relative improvement of the loss function $\mathcal{L}_{RBA}$ is smaller than 0.001.

c) Baseline Settings: We compare RBA with 2 baselines: linear regression, and Bias Alignment (BA). Linear regression is trained over all of the label data without handling process variations. We add BA for reference, whose performance is treated as the upper bound of RBA, because it neglects the dependency between inter- and -intra wafer variations. BA optimizes an additional bias term for each wafer region:

$$\mathbf{y}_{i,j} = \mathbf{X}_{i,j}\mathbf{w} + \mathrm{b}_{i,j} + \epsilon \tag{15}$$

It should be noted that the learned bias term in BA is not transferable to new wafers or new regions, since it requires testing dies under the new process to estimate the new bias term.

A. Effectiveness of V_{min} bias alignment

We aim to showcase the effectiveness of V_{min} bias alignment in the V_{min} prediction task.

a) Experimental Settings: We consider all three types of V_{min} (DC Scan, AC Scan, and MBST), tested at cold and room temperatures. Our dataset comprises 5 wafers tested under these conditions. For each wafer zone, we allocate 75% of the dies for training and the remaining 25% for testing. The methods under consideration include linear regression, BA, and RBA.

b) Results: We report the testing RMSE of V_{min} prediction in each wafer zone, and the average result across the whole testing dataset in Table III for DC Scan V_{min}, Table IV for AC Scan V_{min}, and Table V for MBST V_{min}. The bold number represents the best method, and the underlined number represents the second-best method. In each V_{min} prediction task, both BA and RBA consistently outperform the baseline linear regression. This suggests that the bias alignment technique effectively captures process variations. Notably, the performance gap between BA and RBA is minimal, indicating a weak dependency between inter- and intra-wafer variations.

Additionally, we show the inter- and intra-wafer V_{min} shift estimated by our approach RBA in Table VI and Table VII,

Table VI: Inter-wafer V_{min} shift (mV) from wafer 1 estimated by RBA

Wafer ID	Temp.	DC Scan V_{min}	AC Scan V_{min}	MBST V_{min}
2	-45°C	-18.78	-6.93	-14.41
3	-45°C	-4.32	-4.45	-6.87
4	-45°C	-5.98	5.69	-28.93
5	-45°C	-3.28	2.35	17.68
2	25°C	-9.91	-10.03	-18.75
3	25°C	2.18	-4.10	-8.77
4	25°C	0.62	6.60	-27.99
5	25°C	-0.37	7.48	-18.02

Table VII: Intra-wafer V_{min} shift (mV) from center zone estimated by RBA

Wafer Zone	Temp.	DC Scan V_{min}	AC Scan V_{min}	MBST V_{min}
Inner Donut	-45°C	2.48	4.16	7.35
Outer Donut	-45°C	2.87	6.81	-2.17
Edge	-45°C	1.36	5.65	-1.18
Inner Donut	25°C	2.32	4.22	-1.03
Outer Donut	25°C	3.13	6.95	-2.24
Edge	25°C	-0.35	3.47	-11.15

Table VIII: RBA with different fractions of data for training

Temperature	-45°C		25°C	
Training data frac.	75%	5%	75%	5%
DC Scan V_{min} (mV)	3.75	3.85	4.10	3.82
AC Scan V_{min} (mV)	6.33	6.42	6.43	6.54
MBST V_{min} (mV)	4.91	5.03	4.94	5.18

Table IX: Top 1 linear correlation between class probe features and inter-wafer V_{min} shift estimated by RBA

Test Pattern	DC Scan V_{min}	AC Scan V_{min}	MBST V_{min}
Top 1 Linear Corr.	0.892	0.948	0.953

respectively. The variance of both types of V_{min} shift is substantial and cannot be disregarded. A significant V_{min} shift notably impacts the accuracy of linear regression. For instance, linear regression performs badly for predicting DC Scan V_{min} of wafer 2 at the cold temperature, where a -18.78mV V_{min} shift is estimated by RBA.

B. Data Efficiency and Robustness of RBA

We present the performance of RBA on small training datasets to demonstrate its data efficiency and robustness.

a) Experimental Settings: We use 5% dies in each wafer zone for training, and the rest for testing. All other configurations are the same as those in Section IV-A.

b) Results: The V_{min} prediction accuracy of RBA is listed in Table VIII. While the fraction of training data is reduced from 75% to 5% (around 7 dies in each wafer zone), RBA's accuracy is stable, indicating its superior data efficiency and robustness.

C. Class Probe Features Capturing Inter-Wafer Variation

We demonstrate that the dependency between class probe features and inter-wafer variations estimated by RBA is really high, indicating the motivation to leverage wafer-level class probe features to model wafer-to-wafer variation is plausible.

a) Experimental Settings: Our dataset has 10 wafers whose DC Scan, AC Scan, and MBST V_{min} are tested at the hot temperature. In the first step, we employ RBA on these wafers to collect 9 V_{min} shift terms relative to a base wafer. Subsequently, we correlate these shifts with each class probe feature, reporting the highest absolute value of the Pearson correlation coefficient. A higher coefficient indicates a stronger linear correlation. In the second step, we utilize the V_{min} shifts of 6 wafers to fit a linear model for 2 class probe features and evaluate its testing coefficient of determination (R^2) on the remaining 4 wafers.

b) Results: In Table IX, it is evident that for each V_{min} test pattern, there exists a class probe feature with a correlation coefficient of at least 0.89, indicating (1) the credibility of the inter-wafer V_{min} shift estimated by RBA, and (2) the informativeness of class probe data in modeling V_{min} shift across wafers.

Table X: Coefficient of determination (R^2) of the linear model using 2 class probe features to predict V_{min} shift estimated by RBA

Test Pattern	DC Scan V_{min}	AC Scan V_{min}	MBST V_{min}
R^2	0.509	0.631	0.792

Table X reports the test accuracy of using class probe features to model V_{min} shift. The R^2 score of each V_{min} test pattern is proportion to the Pearson score in Table IX. While MBST V_{min} shift predictors appear promising, we encounter difficulty in obtaining a sufficiently accurate predictor for the DC/AC Scan V_{min} shift. This challenge may stem from the small size of the training dataset, leading to an increased variance. We defer this issue to future research endeavors where a larger pool of tested wafers can be obtained.

Table XI: The testing RMSE (mV) of V_{min} tested at 125°C

Method	DC Scan V_{min}	AC Scan V_{min}	MBST V_{min}
Linear Regression	**7.33**	9.40	14.24
RBA	8.07	**8.98**	**9.18**

D. RBA for V_{min} Prediction of New Wafer

We assess the effectiveness of RBA in predicting V_{min} of new wafers, focusing on addressing inter- and intra-wafer variations.

a) Experimental Settings: We evaluate RBA on the V_{min} prediction task, where V_{min} is tested at 125°C. Following Section IV-C, we use 6 wafers for training and 4 wafers for testing. The baseline model is linear regression.

b) Results: Table XI presents the RMSE of RBA and linear regression. Owing to the limited number of wafers available in our dataset for training the inter-wafer V_{min} shift predictor, RBA and linear regression yield comparable results for DC/AC Scan V_{min} prediction. However, RBA exhibits a significant performance advantage over linear regression in the MBST V_{min} prediction task, highlighting its efficacy in addressing inter- and intra-wafer variations.

V. Conclusion

This paper introduces restricted bias alignment (RBA), a V_{min} prediction framework designed to systematically capture process variations in semiconductor manufacturing. By leveraging class probe features to model inter-wafer variations and utilizing parametric features to estimate intra-wafer variations, RBA offers a comprehensive approach to address the challenges posed by process variations.

Our experiments conducted on an industrial dataset demonstrate the effectiveness of RBA in mitigating the impact of process variations on V_{min} prediction. The results highlight the practical utility and robustness of RBA in real-world semiconductor manufacturing scenarios, underscoring its potential to enhance manufacturing efficiency and reliability.

Acknowledgment

This work is supported by an NXP Long Term University (LTU) grant and the National Science Foundation under Grant No. 1956313.

REFERENCES

[1] W.-C. Lin, C. Chen, C.-H. Hsieh, J. C.-M. Li, E. J.-W. Fang, and S. S.-Y. Hsueh, "Ml-assisted vminbinning with multiple guard bands for low power consumption," in *2022 IEEE International Test Conference (ITC)*, 2022, pp. 213–218.

[2] J. Keane, W. Zhang, and C. H. Kim, "An array-based odometer system for statistically significant circuit aging characterization," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 10, pp. 2374–2385, 2011.

[3] T.-B. Chan, P. Gupta, A. B. Kahng, and L. Lai, "Ddro: A novel performance monitoring methodology based on design-dependent ring oscillators," in *Thirteenth International Symposium on Quality Electronic Design (ISQED)*, 2012, pp. 633–640.

[4] Y.-T. Kuo, W.-C. Lin, C. Chen, C.-H. Hsieh, J. C.-M. Li, E. Jia-Wei Fang, and S. S.-Y. Hsueh, "Minimum operating voltage prediction in production test using accumulative learning," in *2021 IEEE International Test Conference (ITC)*, 2021, pp. 47–52.

[5] Q. Shi, X. Wang, L. Winemberg, and M. M. Tehranipoor, "On-chip sensor selection for effective speed-binning," *Analog Integrated Circuits and Signal Processing*, vol. 88, pp. 369–382, 2016.

[6] J. Chen, J. Zeng, L.-C. Wang, M. Mateja, and J. Rearick, "Predicting multi-core system fmax by data-learning methodology," in *Proceedings of 2010 International Symposium on VLSI Design, Automation and Test*, 2010, pp. 220–223.

[7] W.-C. Lin, C. Chen, C.-H. Hsieh, J. C.-M. Li, E. J.-W. Fang, and S. S.-Y. Hsueh, "Ml-assisted vminbinning with multiple guard bands for low power consumption," in *2022 IEEE International Test Conference (ITC)*, 2022, pp. 213–218.

[8] Y. Yin, R. Chen, C. He, and P. Li, "Domain-specific machine learning based minimum operating voltage prediction using on-chip monitor data," in *2023 IEEE International Test Conference (ITC)*, 2023, pp. 99–104.

[9] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[10] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "Catboost: unbiased boosting with categorical features," vol. 31, 2018.

[11] Y. Yin, X. Wang, R. Chen, C. He, and P. Li, "Reliable interval prediction of minimum operating voltage based on on-chip monitors via conformalized quantile regression," in *2024 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2024, pp. 1–6.

[12] M. A. Hall, "Correlation-based feature selection for machine learning," Ph.D. dissertation, The University of Waikato, 1999.