

# Harnessing Fabrication Process Signature for Predicting Yield Across Designs

Ali Ahmadi\*, Haralampos-G. Stratigopoulos†, Amit Nahar‡, Bob Orr‡, Michael Pas‡ and Yiorgos Makris\*

\*Department of Electrical Engineering, The University of Texas at Dallas, Richardson, TX 75080

†Sorbonne Universités, UPMC Univ. Paris 06, CNRS, LIP6, 4 place Jussieu, 75005, Paris, France

‡Texas Instruments Inc., 12500 TI Boulevard, MS 8741, Dallas, TX 75243

**Abstract**—Yield estimation is an indispensable piece of information at the onset of high-volume manufacturing (HVM) of a device. The increasing demand for faster time-to-market and for designs with growing quality requirements and complexity, requires a quick and successful yield estimation prior to HVM. Prior to commencing HVM, a few early silicon wafers are typically produced and subjected to thorough characterization. One of the objectives of such characterization is yield estimation with better accuracy than what pre-silicon Monte Carlo simulation may offer. In this work, we propose predicting yield of a device using information from a similar previous-generation device, which is manufactured in the same technology node and in the same fabrication facility. For this purpose, we rely on the Bayesian Model Fusion (BMF) technique. The effectiveness of the proposed methodology is evaluated using sizable industrial data from two RF devices in a 65nm technology.

## I. INTRODUCTION

The trend nowadays is towards mixed-signal Systems-on-Chip (SoCs), wherein analog and RF circuits are integrated together with the digital processor, memory, etc. Towards this goal, analog and RF devices are now designed in advance technology nodes and, as a result, they suffer from increased process variations which may lead to significant yield loss. Therefore, accurate and fast prediction of yield of a new device is an indispensable piece of information during production, in order to identify and quickly resolve any issues that may jeopardize production ramp-up. To this end, significant effort has been invested in improving and speeding up Monte Carlo-based yield estimation [1]–[3].

In the rapidly growing and dynamically changing consumer electronics market, time-to-market is a crucial factor in investment return. The semiconductor industry often reuses an existing device and implements slight modifications and enhancements to develop the next-generation device so as to respond to market demands in a reasonable time.

In this work, we introduce a methodology to predict yield of a device which is planned to be produced in HVM in a fabrication facility, by borrowing information from a previous-generation device that is currently being produced or was produced in the past in HVM in the same fabrication facility.

To accomplish this, we rely on two facts. First, two devices fabricated in the same technology node and in the same fabrication facility experience very similar process variations. Therefore, they share similar e-test distributions, where by the term e-test we refer to electrical measurements which are typically performed using process control monitors (PCMs) that are included in the wafer scribe lines in select locations across

the wafer. Second, since the new-generation device has slight modifications and design improvements as compared to the previous-generation device, both devices exhibit a very similar performance deviation pattern due to process variations.

The proposed methodology relies on modeling yield of a wafer as a function of its e-tests. This enables us to predict yield of a wafer solely based on its e-tests. Such a prediction model can be learned reliably for the previous-generation device thanks to the large volume of data that is available. In this work, we deal with the problem of learning such a prediction model for the new-generation device during the characterization phase, where only a few early wafers with the new-generation device are available. Thereafter, the HVM yield of the new-generation device can be predicted by considering the available e-test profile of the previous-generation device.

To accomplish this, we employ the BMF learning procedure which aims at effectively refining and adapting the prediction model for the new-generation device by incorporating, in an intelligent manner, prior knowledge from the previous-generation device. BMF is a very powerful technique which has been used successfully for model improvement in various contexts in the past, including pre-silicon validation, post-manufacturing tuning, bit error rate estimation, alternate test, and production migration [4]–[8].

The proposed BMF learning procedure is compared with three other more straightforward HVM yield prediction methods.

## II. YIELD/E-TEST CORRELATION

Let us consider device A that is currently being produced in HVM in a specific fabrication facility. Let us also assume that we have at hand the e-test measurements from  $w_A$  wafers that contain device A and the probe-tests from all devices contained in each of these wafers, where by the term probe-tests we refer to electrical measurements performed to derive the performances of the device. Formally, let  $ET_A^i = [ET_{A,1}^i, \dots, ET_{A,l}^i]$  denote the  $l$ -dimensional e-test measurement pattern of the  $i$ -th wafer, where  $ET_{A,k}^i$  denotes the  $k$ -th e-test measurement. By knowing the specification limits for all probe-tests, we can compute yield of the  $i$ -th wafer, denoted by  $y_A^i$ , as the percentage of devices in the  $i$ -th wafer that pass all probe-test specification limits. Thus, information from device A includes

$$\text{wafer}_A^i = [ET_A^i, y_A^i], \quad i = 1, \dots, w_A. \quad (1)$$

Using the training data in (1), we can learn the correlation between yield and e-test measurements of a wafer using a regression function

$$\hat{y}^i \approx f_A(\mathbf{ET}^i). \quad (2)$$

Once the regression function is learned, we can use it to predict the yield  $\hat{y}^i$  for future wafers containing device A, i.e. for  $i > w_A$ , based on their e-test measurements.

### III. YIELD PREDICTION ACROSS DESIGNS

Let us now consider that device B is a next generation of device A with slight modifications and improvements and that device B is planned to be produced in HVM in the same technology node and fabrication facility that device A is currently being or was produced. Let us assume that we have at hand the e-test measurements from the first  $w_B$  wafers that contain device B and the probe-tests from all devices contained in each of these wafers. Following similar notation as in Section II, information from device B includes

$$\text{wafer}_B^i = [\mathbf{ET}_B^i, y_B^i], \quad i = 1, \dots, w_B. \quad (3)$$

We are interested in using the limited data in (3) to accurately predict HVM yield of device B.

#### A. Averaging

A simple and straightforward approach is to compute the average yield of the  $w_B$  early wafers and use it as an estimation of HVM yield of device B

$$\bar{y}_B = \frac{1}{w_B} \sum_{i=1}^{w_B} y_B^i. \quad (4)$$

#### B. Early learning

Another approach is to use the data in (3) as a training set and learn a regression model to express yield as a function of the e-tests for device B

$$\hat{y}^i \approx f_B(\mathbf{ET}^i). \quad (5)$$

The HVM yield of device B can be predicted by employing the e-test profile of device A, since it is very similar to that of device B

$$\bar{y}_B = \frac{1}{w_A} \sum_{i=1}^{w_A} f_B(\mathbf{ET}_A^i). \quad (6)$$

#### C. Naive mixing of data

Another approach is to naively mix data in (1) and (3), use the combined data as a training set, and learn a regression model to express yield as a function of the e-tests

$$\hat{y}^i = f_{AB}(\mathbf{ET}^i). \quad (7)$$

The HVM yield of device B can be predicted as

$$\bar{y}_B = \frac{1}{w_A} \sum_{i=1}^{w_A} f_{AB}(\mathbf{ET}_A^i). \quad (8)$$

#### D. Bayesian Model Fusion

The BMF approach is similar to early learning, but the training procedure leverages information from device A in an intelligent manner. In particular, for devices A and B we assume regression models

$$f_A(\mathbf{ET}^i) = \sum_{m=1}^M a_{A,m} \cdot b_m(\mathbf{ET}^i) \quad (9)$$

and

$$f_{B,BMF}(\mathbf{ET}^i) = \sum_{m=1}^M a_{B,m} \cdot b_m(\mathbf{ET}^i), \quad (10)$$

respectively. These regression models are based on  $M$  basis functions, where  $b_m$  is the  $m$ -th basis function, and  $a_{A,m}$  and  $a_{B,m}$  correspond to the coefficient of the  $m$ -th basis function for device A and B, respectively. The coefficients  $\mathbf{a}_A = [a_{A,1}, \dots, a_{A,M}]$  of regression model  $f_A$  can be learned accurately based on the rich dataset in (1). The coefficients  $\mathbf{a}_B = [a_{B,1}, \dots, a_{B,M}]$  of regression model  $f_B$  are learned by maximizing the posterior distribution

$$\max_{\mathbf{a}_B} \text{pdf}(\mathbf{a}_B | \text{wafer}_B), \quad (11)$$

where  $\text{pdf}(\mathbf{a}_B | \text{wafer}_B) \propto \text{pdf}(\mathbf{a}_B) \text{pdf}(\text{wafer}_B | \mathbf{a}_B)$ ,  $\text{pdf}(\mathbf{a}_B)$  is the *prior* distribution,  $\text{pdf}(\text{wafer}_B | \mathbf{a}_B)$  is the *likelihood function*, and  $\text{wafer}_B = [\text{wafer}_B^1, \dots, \text{wafer}_B^{w_B}]$ . In this way, we maximize the “agreement” of the selected coefficients with the limited observed data in (3). An expression for the *prior* distribution is developed by involving the prior knowledge from device A, whereas an expression for the *likelihood function* is developed by using the data in (3). Due to the lack of space, the interested reader is referred to [7], [8] for an in-depth discussion on the learning procedure formulation based on BMF.

The HVM yield of device B can, then, be predicted as

$$\bar{y}_B = \frac{1}{w_A} \sum_{i=1}^{w_A} f_{B,BMF}(\mathbf{ET}_A^i). \quad (12)$$

### IV. EXPERIMENTAL RESULTS

#### A. Data set and objectives

We use actual production data from two RF devices fabricated in a 65nm technology in the same fabrication facility by Texas Instruments<sup>1</sup>. We will refer to these devices as device A and device B, following the terminology in the rest of the paper. The dataset for device A includes 54 e-tests obtained on 9 e-test sites and 168 probe-tests for a total of 1800 wafers with approximately 1500 die per wafer. The dataset for device B includes the same 54 e-tests and 200 probe-tests for a total of 1000 wafers with approximately 1500 die per wafer. The e-test signature of a given wafer is computed as the mean and standard deviation of each e-test across the e-test sites, which leads to an e-test signature with 108 features. Along with the data, we are also provided with the specification limits for each probe-test, hence we can compute the actual yield of each wafer.

<sup>1</sup>Details regarding the devices cannot be released due to an NDA under which this data has been provided to us.

Using this dataset, we seek to:

- Quantify the existence of a correlation between yield of a wafer and its e-test signature, which enables precise prediction of yield of a wafer solely based on its e-test signature.
- Confirm that accurate prediction of HVM yield of a device entirely based on data from a few engineering wafers is not feasible.
- Evaluate and compare the different HVM yield prediction methods described in Section III.

The regression functions in the different HVM yield prediction methods in Section III are learned using Multivariate Adaptive Regression Splines (MARS) [9].

### B. Yield/e-test correlation

The accuracy of predicting wafer yield from e-tests is studied independently for both devices A and B by employing the complete data sets in (1) and (3). We learn and assess the generalization of the regression models in (2) and (5) by using 5-fold cross-validation. Specifically, a data set is divided into 5 folds, where 4 folds are used for training and the remaining fold for validation. The procedure is repeated such that all folds are left out as a validation set and in the end we report the average prediction error. We employ the absolute prediction error defined as

$$\delta_i = |\hat{y}^i - y^i|, \quad (13)$$

where  $\hat{y}^i$  and  $y^i$  are the predicted and the actual yield of the  $i$ -th wafer, respectively.

Figures 1(a) and 1(b) illustrate the correlation between yield and e-tests for devices A and B, respectively. In each histogram, the x-axis represents absolute prediction error in % and the y-axis represents wafers in the validation set in %. Each bin of the histogram shows the percentage of wafers in the validation set for the corresponding prediction error range. For example, regarding device A, the yield of about 37% of wafers in the validation set is predicted with an error in the range 0-0.5%. Figures 1(a) and 1(b) also illustrate with vertical lines the average  $\delta_{avg}$  and maximum  $\delta_{max}$  absolute prediction errors in % across the validation set. As it can be seen, for both devices, the yield can be predicted with an average prediction error close to 1% and a maximum prediction error that does not exceed 5%. This corroborates our conjecture that the correlation between e-tests and yield is strong, which allows us to predict yield from e-tests using a regression function, provided that the training set is rich and representative of HVM.

### C. Yield prediction across designs

In order to demonstrate and compare the HVM yield prediction methods proposed in Section III, we performed the following experiment. We assume access to the entire dataset of device A, which constitutes the training data in (1) with  $w_A = 1800$ . For device B, we assume that we have available only a subset of wafers, in particular wafers that come from the first two lots. We vary  $w_B$  in the range [10, 50] and we

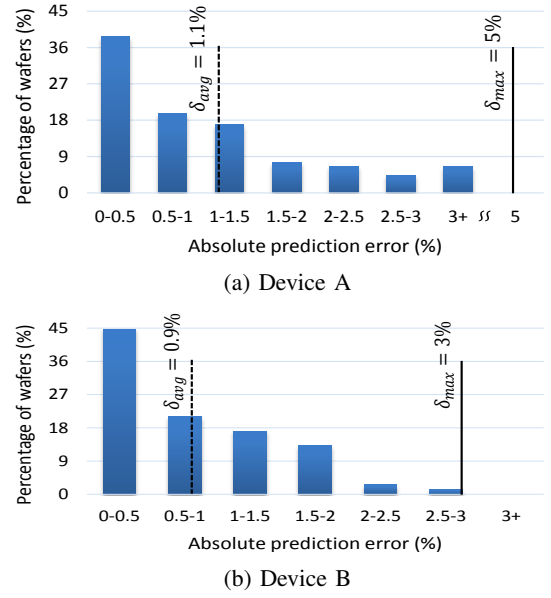


Fig. 1: Error in predicting yield from e-tests.

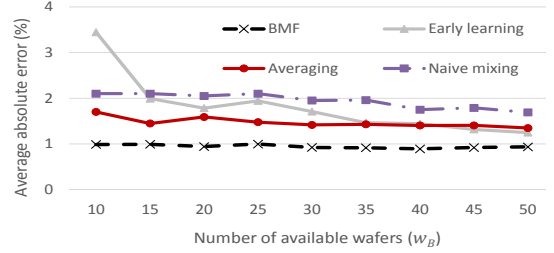


Fig. 2: Average HVM yield prediction error for device B from a few early wafers.

employ the methods proposed in Section III to predict the HVM yield. We report the average absolute prediction error expressed as

$$\frac{1}{w'_B} \sum_{i=1}^{w'_B} |\bar{y}_B - y^i|, \quad (14)$$

where  $w'_B$  denotes the size of the validation set defined as the available wafers for device B excluding the  $w_B$  wafers, that is,  $w'_B = 1000 - w_B$ ,  $\bar{y}_B$  is the HVM yield prediction, and  $y^i$  is the actual yield of the  $i$ -th wafer.

The accuracy of the different HVM yield prediction methods proposed in Section III is presented in Figure 2. The curves show the average absolute prediction error as a function of  $w_B$ . As it can be seen, the BMF outperforms the other straightforward methods regardless of the size of the training set. It shows a remarkably stable behavior, maintaining steady HVM yield prediction error even when the training set size is as small as 10 wafers. This shows that the BMF method, by statistically fusing prior knowledge from device A, is capable of providing a very accurate HVM yield prediction model for device B, based on only a few early fabricated wafers of device B. Therefore, BMF can be used for a fast and precise forecasting of HVM yield from a few early wafers in the HVM, without having to wait until a large volume of data

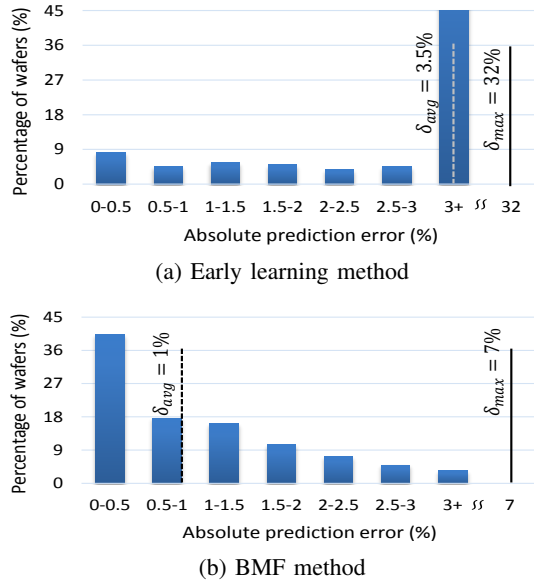


Fig. 3: Yield prediction error for device B when  $w_b = 10$ .

is collected. The second best method is the averaging method. The stable behavior implies that the yield of the wafers in the first two lots that are included in the training set is very similar. It is outperformed by the BMF method since the wafers in the first two lots are not very representative of HVM. The early learning method strongly depends on the size of the training set. The prediction error is low for large  $w_b$  and exponentially increases as  $w_b$  becomes smaller. This is anticipated since the information content of the training set is weakened, becoming biased and non-representative of HVM, and the regression model is unable to extrapolate towards the tails of distribution, resulting in large prediction error. The naive mixing of data method is outperformed by all other methods for  $w_b > 15$  and surpasses only the early learning method for  $w_b = 10$ . The fact that the accuracy of this method is inferior implies that the data from devices A and B do not exhibit strong similarity and that the rich data from device A overshadow the limited data from device B.

To gain better insight, we consider  $w_b = 10$  and we illustrate in Figure 3 the distribution of absolute prediction error for all wafers in the validation set for the BMF and early learning methods, which have the best and worst predictions for this value of  $w_b$ . The absolute prediction error is calculated as in (13). As in Figure 1, in each histogram, the x-axis represents absolute prediction error in % and the y-axis represents wafers in the validation set in %. Each bin of the histogram shows the percentage of wafers in the validation set for the corresponding prediction error range. As it can be seen, for the BMF method the histogram is skewed to the left, showing that the yield of the majority of the wafers is predicted accurately with average and maximum errors 1% and 7%, respectively, whereas for the early learning method the histogram is skewed to the right, showing that the yield of about half of the wafers is predicted with an error of 3.5% and that the maximum error reaches 32%.

Finally, regarding the BMF method, by comparing Figures 3(b) and 1(b), we observe that information from as few as 10 fabricated wafers of device B suffices to reduce HVM yield prediction error to the quality of prediction that employs a large HVM population of 800 wafers.

## V. CONCLUSION

We discussed methods to accurately predict HVM yield of a device from a few early silicon wafers assuming availability of data from a previous-generation device. The set of methods includes three rather straightforward methods and a new more sophisticated method based on BMF. As demonstrated using a large dataset from two 65nm devices from Texas Instruments, the BMF method shows a very stable performance and outperforms the straightforward methods, since it can intelligently combine data from the new-generation and previous-generation devices. By using only 10 wafers from the first two lots and including in the analysis prior information from a previous-generation device, the BMF method is capable of predicting HVM yield within 1% of error.

## VI. ACKNOWLEDGMENT

This research has been partially supported by the Semiconductor Research Corporation (SRC) Task 1836.131.

## REFERENCES

- [1] C. M. Kurker, J. J. Paulos, R. S. Gyurcsik, and J.-C. Lu, "Hierarchical yield estimation of large Analog integrated circuits," *IEEE Journal of Solid-State Circuits*, vol. 28, no. 3, pp. 203–209, 1993.
- [2] B. Liu, F. V. Fernández, and G. G. E. Gielen, "Efficient and accurate statistical Analog yield optimization and variation-aware circuit sizing based on computational intelligence techniques," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 30, no. 6, pp. 793–805, 2011.
- [3] F. Gong, H. Yu, Y. Shi, and L. He, "Variability-aware parametric yield estimation for Analog/Mixed-signal circuits: Concepts, algorithms, and challenges," *IEEE Design & Test*, vol. 31, no. 4, pp. 6–15, 2014.
- [4] X. Li, W. Zhang, F. Wang, S. Sun, and C. Gu, "Efficient parametric yield estimation of Analog/Mixed-signal circuits via Bayesian model fusion," in *Proc. IEEE/ACM International Conference on Computer-Aided Design*, 2012, pp. 627–634.
- [5] F. Wang, W. Zhang, S. Sun, X. Li, and C. Gu, "Bayesian model fusion: large-scale performance modeling of Analog and Mixed-signal circuits by reusing early-stage data," in *Proc. IEEE/ACM Design Automation Conference*, 2013, pp. 59–64.
- [6] C. Gu, E. Chiprout, and X. Li, "Efficient moment estimation with extremely small sample size via Bayesian inference for Analog/Mixed-signal validation," in *Proc. IEEE/ACM Design Automation Conference*, 2013, pp. 1–7.
- [7] J. Liaperdos, H.-G. Stratigopoulos, L. Abdallah, Y. Tsiatouhas, A. Arapoyanni, and X. Li, "Fast deployment of alternate Analog test using Bayesian model fusion," in *Proc. Design, Automation & Test in Europe Conference*, 2015, pp. 1030–1035.
- [8] A. Ahmadi, H.-G. Stratigopoulos, A. Nahar, B. Orr, M. Pas, and Y. Makris, "Yield forecasting in Fab-to-Fab production migration based on Bayesian model fusion," in *Proc. IEEE/ACM International Conference on Computer-Aided Design*, 2015, pp. 9–14.
- [9] J. H. Friedman, "Multivariate adaptive regression splines," *The Annals of Statistics*, vol. 19, no. 1, pp. 1–67, 1991.