Early Prediction of Product Performance and Yield Via Technology Benchmark

Choongyeun Cho, Daeik D. Kim
IBM Semiconductor R&D Center
Hopewell Junction, NY
{cycho,dkim}@us.ibm.com

Jonghae Kim
Qualcomm Inc.
San Diego, CA
jonghaek@qualcomm.com

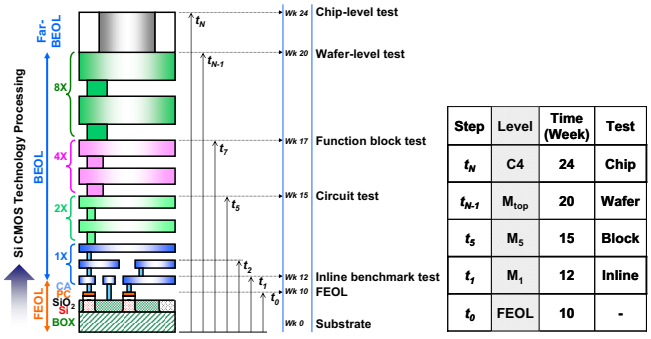
Daihyun Lim
MIT MTL
Cambridge, MA 02139
daihyun@mit.edu

Sangyeun Cho
Univ. of Pittsburgh
Pittsburgh, PA 15260
cho@cs.pitt.edu

***Abstract*—This paper presents a practical method to estimate IC product performance and parametric yield solely from a well-chosen set of existing electrical measurements intended for technology monitoring at an early stage of manufacturing. We demonstrate that the components of mmWave PLL and product-like logic performance in a 65nm SOI CMOS technology are predicted within a 5% RMS error relative to mean.**

I. Introduction

In a deep sub-micron technology, process variability has become a predominant limiter of performance and yield of IC products [1]. At the same time, it becomes increasingly more difficult to model and predict it accurately, partly due to the complicated nature of process variations [2]. While one can use a simplified analytical model or Monte-Carlo circuit simulation to directly estimate a product circuit's performance, this often leads to inaccuracies because the process variation is amplified by specific parameters, rather than the overall parameter distributions. The nonlinearity becomes especially pronounced in modern deep sub-micron devices whose variation is dominated by the device parameters and parasitic components.



Step	Level	Time (Week)	Test
t_N	C4	24	Chip
t_{N-1}	M_{top}	20	Wafer
t_5	M_5	15	Block
t_1	M_1	12	Inline
t_0	FEOL	10	-

Fig. 1. Schedule for CMOS technology development and testing. The motivation of this work is that the product performance is typically tested only after C4 level is completed, but in-line benchmark measurements at M1, if carefully planned and selected, carry relevant information to reliably predict the product performance.

Fig. 1 shows a typical schedule for CMOS technology development and testing at different levels. Typical elapsed time for key technology process steps (FEOL, M1, M4, top metal and C4) are arranged in the table. The purpose of this work is to reliably estimate the performance and yield of a product at an early stage of IC manufacturing, especially at the first metal level (M1).

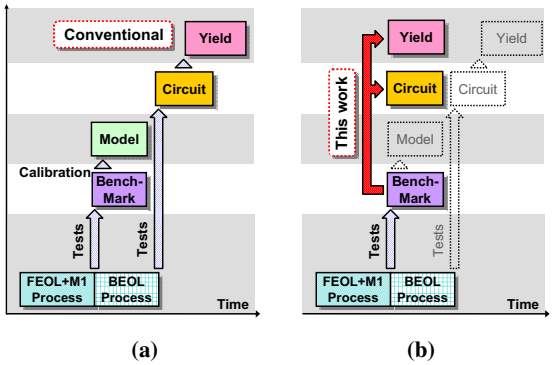


Fig. 2. (a) A typical IC development flow in terms of process, benchmark, model, and circuit/product test. (b) Flow of the proposed early prediction of product performance and yield based on manufacturing in-line benchmark structures. Note that through prediction of product performance/yield from benchmark, the time and cost for testing/qualification reduce significantly.

Often, accurate product performance and yield information becomes available only after dicing and packaging. Fig. 2(a) illustrates a typical IC development time line that involves different development components – process, benchmark, model, and a product circuit. A model and benchmarking are strongly tied together at all times through many rounds of model calibration via model-to-hardware correlation (MHC). Circuit performance and yield are conventionally measured and qualified after the back-end-of-line (BEOL) processes are completed. This long cycle of production to yield hinders fast yield learning and product ramp-up, adding greatly to the product development and manufacturing cost. For early estimation, we utilize a well-chosen set of existing electrical measurements from *manufacturing in-line benchmark structures* (MIBS). The MIBS in this paper collectively refer to assorted test structures that are measured in a manufacturing line using a standard parametric tester for the purpose of defect diagnosis, DC device characterization, and MHC [3]. Typical MIBS include test structures for MOSFET devices having different sizes and layouts, ring oscillators (ROs), resistance/capacitance structures. MIBS data are typically collected

in an automated fashion for all or most wafers at a lower metalization level, usually at M1. A rationale for utilizing measured MIBS data is that the process-induced variability in key technology parameters, such as threshold voltage (V_{th}), oxide thickness (T_{ox}), and gate length (L_{poly}) dominates the variation of circuit performance. We prudently select a set of MIBS at early stage of manufacturing (e.g. M1) such that the variability of these key physical parameters is embedded therein. The proposed methodology is validated using an ensemble of MIBS data collected from test structures built with a 65nm SOI CMOS technology.

While a prior work used MIBS data to identify the most correlated device characteristics with respect to circuit performance [4], MIBS have seldom been exploited for the purpose beyond what they are intended. FET devices and ROs are popularly used to characterize CMOS circuit performance and variability [5]. However, there has been little practice for a statistical approach to characterize and qualify circuit product's performance using a collection of heterogeneous MIBS measurements, mainly due to the increasing complexity and nonlinearity of the technology-product relationship.

We anticipate that the method proposed in this work will have large, practical impact on product yield-learning acceleration and testing time/cost reduction. An early prediction of a product circuit performance and yield based on MIBS data allows rapid yield learning, thus saving development cost and time (see Fig. 2(b)). Also to design community, the mapping from device parameters to product performance metric can provide pertinent feedback in order to modify circuit more tolerant to particular process variation source.

The rest of this paper is organized as follows. In Section II, we describe the proposed estimation method in detail. In Section III, our estimation method is evaluated using two test vehicles, an RF frequency divider and an RO, which are representative of important building blocks for clocking and logic product, respectively. A summary will follow, in the conclusion.

II. Proposed Performance/Yield Prediction

In this section, we briefly introduce a multivariate statistical technique to estimate circuit performance and yield from MIBS measurements. The method is based on feature extraction and nonlinear estimation, widely used in semiconductor manufacturing industry [6], [7]. Feature vectors are extracted in order to reduce the dimensionality of input data. A nonlinear estimation uses an ensemble-based training of a neural network. Neural nets are used to learn and compute functions for which the relationships between input and output vectors are unknown or computationally complicated which is generally true for product circuit parameters with respect to device-level parameters.

A. Projected Principal Component Analysis

The number of MIBS parameters monitored on a regular basis can be on the order of thousands or tens of thousands, and it is not feasible to stably and reliably train an estimator based on all raw MIBS measurements. A variant of principal component analysis (PCA) is employed to extract features in all MIBS data. The MIBS measurements are projected onto the subspace of circuit performance parameter(s) *while preserving the most information relevant to the circuit parameters*. Hence, this transform is called *projected* PCA.

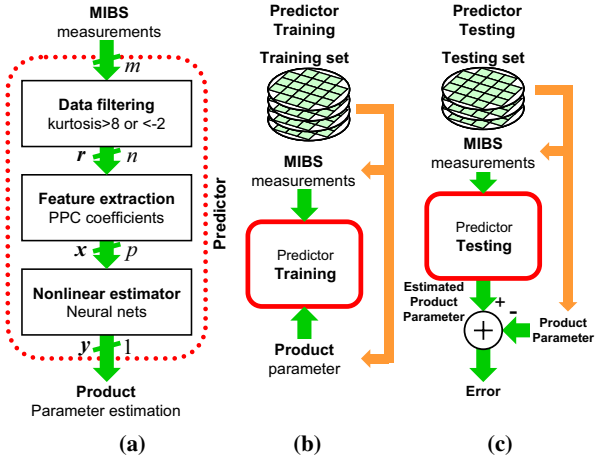


Fig. 3. (a) Flow chart of the proposed nonlinear estimator for circuit parameters. (b) Training of the estimator. (c) Testing scheme.

B. Performance/Yield Prediction Algorithm

Fig. 3(a) illustrates the proposed nonlinear estimation algorithm. First, the input MIBS data of m parameters is screened to include only usable data of size n. In this work, we employed a simple fourth-momentum test to filter out data with abnormal distribution. Second, the resulting data is compressed to a manageable number of parameters, p. Projected PCA is used to extract feature vectors. The selection of how many PC's to retain (p) for the subsequent steps is based on a trade-off between the accuracy of estimation and the stability of the estimation training (fitting). Third, feature vectors are put into a nonlinear estimator to predict circuit performance parameter(s).

Fig. 3(b) shows how the estimation algorithm is trained. Using a pair of *actual* MIBS measurements (input) and product circuit parameter (output), the weights of the neural networks are adjusted. Fig. 3(c) illustrates how the algorithm is tested; the error will be calculated based on predicted and actual circuit parameters. We use percentage RMS error as a metric for estimation error.

Yield estimation (binary detection) is an extension of circuit parameter prediction (continuous estimation). After estimating circuit performance parameter, a yield is calculated as the fraction of samples that meet pre-defined specifications. For example, this method accommodates design specifications such as maximum gate delay and maximum current (power) in the case of an RO.

TABLE I
CATEGORY OF M1 MIBS USED IN THIS EXPERIMENT.

Test structure category	# of parameters (before screening)	# of parameters (after screening)
FET	1,988	759
Ring oscillator	248	83
SRAM	398	159
Capacitance	222	108
Total	2,856	1,109

III. PRODUCT PERFORMANCE/YIELD PREDICTION

For our experiments, 65nm SOI CMOS technology data is used. Table I categorizes the MIBS used in our experiments. All MIBS used in the experiments are at the first metal level (M1).

The number of parameters at each stage in Fig. 3 is: m=2,856 (the number of all MIBS parameters), n=1,109 (the number of MIBS parameters after screening), and $p = 10$ (the number of PPC's used for the estimation). 390 chips (from 10 wafers and 39 chips per wafer) were used for fitting a linear regression, or training the neural nets, and 2 other wafers (78 chips) were used for the testing purpose. The neural net configuration used in these experiments include one hidden layer with 4 nodes ($k = 4$). The number of training samples are more than 5 times the degrees of freedom in the neural net estimator, exceeding the general rule of thumb for neural net training set size.

We used two common types of product circuits for the validation of the proposed methodology: a CML frequency divider operating at up to 90GHz, and a static RO operating at around 1GHz. They in general represent essential building blocks for clocking and logic products. Fig. 4 illustrates approximate locations of the frequency divider, RO and MIBS used in this experiment. The RF divider and RO's are adjacent within a die, and MIBS are clustered in two large columns.

A. mmWave System Clock Component

A typical phase-locked loop (PLL) block uses a loop structure to lock a free-running voltage-controlled oscillator (VCO) to a desired frequency. A primary frequency divider is one of the key PLL components because it must reliably divide VCO operating frequency into a desired lowered frequency as mmWave circuit [8]. Self-oscillation frequency (f_{SO}) is an important figure-of-merit for a frequency divider because it is closely linked to a maximum dividable bandwidth and a sensitivity curve, and is also practical for measurement.

The percentage error (RMS error normalized to mean value) for f_{SO} is arranged in Table II for linear fit and neural net based nonlinear estimator. Estimation errors are approximately 5% of its mean for all estimation schemes. A yield, here, is defined as percentage of chips that meet the product specification: f_{SO} is higher than 34GHz and active current is lower than 27mA. The neural net estimator is slightly more accurate than linear fit due to its flexibility and robustness to highly nonlinear relationship of input (device-level parameters) and output (circuit-level parameters). Neural net is, however,

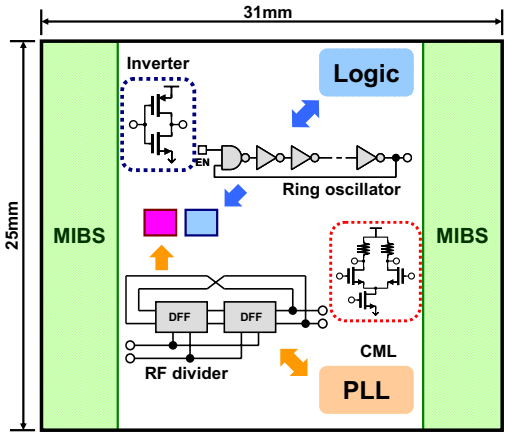


Fig. 4. Die size and approximate locations of product circuits (test vehicles for the validation of the proposed method) and MIBS.

superior for the yield estimation where nonlinearity is more severe. Thus, neural net is preferred overall for estimation of both performance metric and yield.

B. Microprocessor Logic Product

ROs are commonly used as logic-type test vehicles for MHC and variability monitoring. Gate delay (τ) is known to be representative of product environment [9], thus was selected as a subject of the proposed estimation. In this experiment, τ of 101-stage inverter-based ROs was estimated using the same MIBS data as in the previous subsection. (83 RO-related MIBS parameters were excluded in experiment for the fairness of prediction.) Table III presents the normalized RMS estimation errors. A yield for this case is defined as ratio of chips satisfying the specification: gate delay is less than 4.1ps and active current is lower than 2.2mA.

Using mostly FET characteristics, gate delay was estimated within a 4% error for all estimation schemes. The estimation result for yield shows a similar trend as that of the frequency divider: Neural network outperforms linear and second-order regression fits because yield is a highly nonlinear function of design parameters. There are a number of factors contributing to the estimation error: (1) BEOL process variability (above M1) is excluded for estimation since MIBS only up to M1 level were taken into account; (2) Intra-die variation between test vehicles and MIBS was not considered. In spite of these factors, the prediction accuracy obtained with our methodology is reasonably high.

C. Validation of Prediction Robustness

The proposed method uses only the statistics of hardware measurements without considering an underlying physical model. Validity of the estimation errors in Tables II, III can be questioned, especially if training and testing sets are *inbred* in nature, or deliberately selected to minimize estimation error. In this subsection, the robustness of the presented estimation method is evaluated via different combinations of training

TABLE II
ESTIMATION ERROR FOR RF FREQUENCY DIVIDER (SEE SECTION III-A).

Parameter type	Mean	Std dev	Linear fit %err	Neural net %err
Self-oscillation frequency (f_{SO})	34.6GHz	2.9GHz (8.3%)	5.1%	4.6%
I_{dd}	24.0mA	3.6mA (15.1%)	10.1%	10.0%
Yield	27.5%	–	19.3%	17.7%

TABLE III
ESTIMATION ERROR FOR RING OSCILLATOR (SEE SECTION III-B).

Parameter type	Mean	Std dev	Linear fit %err	Neural net %err
RO gate delay (τ)	4.3ps	0.37ps (8.7%)	3.8%	3.5%
$I_{A,RO}$	2.5mA	0.28mA (11.1%)	5.5%	5.1%
Yield	25.6%	–	15.6%	13.1%

TABLE IV
f_{SO} ESTIMATION ERROR FOR FOUR TESTING/TRAINING SCENARIOS.

Training set	Testing set	Percentage error
Slow	Slow	4.5%
Slow	Fast	5.6%
Fast	Slow	5.5%
Fast	Fast	4.9%

set and testing set. Using f_{SO} as a criterion, 12 available wafers are divided into two groups of equal size – faster and slower wafers, relative to the threshold f_{SO}=35GHz. Due to the limited sample size, a linear regression fit was used in this experiment. Better estimation accuracy is expected with a neural net nonlinear estimator if a larger data set is available.

The resulting estimation errors are shown in Table IV. Percentage errors for all training/testing combinations are in a narrow range from 4.5% to 5.6%, showing that the proposed estimation method is robust and can predict circuit performance and yield accurately even when significant wafer variations are present.

D. Feedback to Circuit Design and Technology

Previously, we have validated the proposed methodology via two test circuits, an RF frequency divider and an RO. The mapping from MIBS parameters to product performance metric (learned from training of an estimator) can lead to useful insight as to what device characteristics are most sensitive to product performance, and to what degree. In addition to conventional circuit simulation and analysis, a designer can leverage this information to make circuit more robust to process variation, and to make an educated trade-off between design parameters. For example, the three most dominating device characteristics with relation to the RO delay are found to be FET on / off currents and threshold voltage (obtained by a linear estimation in Section III-B) with the relationship:

$$\tau \propto I_{on} + 0.817 I_{off} + 0.765 V_{th} + \epsilon \qquad (1)$$

where each parameter is normalized to zero-mean and unit-variance.

IV. CONCLUSION

An efficient statistical method was presented to predict circuit performance and yield based on MIBS intended for technology monitoring and device characterization. It predicts the mmWave frequency divider performance within 5% error, and the RO gate delay within 4% error. The significances of this work are: (1) For the first time, we verified the potential of utilizing existing up-to-M1 device-level measurements to directly predict product circuit's performance and yield without assuming any physical model. (2) In-line measurements are available early in the fabrication cycle, *e.g.*, at M1 level. Hence, the performance/yield of a complex, customized RF circuit can be predicted while still in a manufacturing line, allowing significant reduction of the cost and time for product circuit testing. (3) The mapping from device parameters to product performance metric is useful to designers to make a circuit more robust to process variation.

ACKNOWLEDGEMENT

The authors thank IBM SRDC engineers P. O'Neil, C. Schnabel, K. Warren, K. Ginn, and G. Patton for their support, and D. Boning at MIT for valuable comments.

REFERENCES

[1] S. Nassif, "Modeling and analysis of manufacturing variations," in *Proc. IEEE CICC*, 2001, pp. 223–228.

[2] S. Samaan, "The impact of device parameter variations on the frequency and performance of VLSI chips," in *Proc. IEEE/ACM ICCAD*, 2004, pp. 343–346.

[3] M. Ketchen, M. Bhushan, and D. Pearson, "High speed test structures for in-line process monitoring and model calibration," in *Proc. IEEE ICMTS*, 2005, pp. 33–38.

[4] C. Cho, D. Kim, J. Kim, J.-O. Plouchart, and R. Trzcinski, "Statistical framework for technology-model-product co-design and convergence," in *Proc. ACM/IEEE DAC*, San Diego, CA, 2007, pp. 503–508.

[5] M. Ketchen and M. Bhushan, "Product-representative "at speed" test structures for CMOS characterization," *IBM J. Res. & Dev.*, vol. 90, no. 4/5, pp. 451–468, Jul/Sep 2006.

[6] D. White, D. Boning, S. Butler, and G. Barna, "Spatial characterization of wafer state using principal component analysis of optical emission spectra in plasma etch," *IEEE Trans. Semiconduct. Manufacf.*, vol. 10, no. 10, pp. 52–61, Feb 1997.

[7] F. Chen and S. Liu, "A neural-network approach to recognize defect spatial pattern in semiconductor fabrication," *IEEE Trans. Semiconduct. Manufacf.*, vol. 13, no. 3, pp. 366–373, Aug 2000.

[8] D. Lim, J. Kim, J.-O. Plouchart, C. Cho, D. Kim, R. Trzcinski, and D. Boning, "Performance variability of a 90GHz static CML frequency divider in 65nm SOI CMOS technology," in *IEEE ISSCC Dig. Tech. Papers*, 2007, pp. 542–621.

[9] M. Bhushan, M. Ketchen, S. Polonsky, and A. Gattiker, "Ring oscillator based technique for measuring variability statistics," in *Proc. IEEE ICMTS*, 2006, pp. 87–92.