Correlating Inline Data With Final Test Outcomes in Analog/RF Devices

Nathan Kupp*, Mustapha Slamani†, and Yiorgos Makris*
*Department of Electrical Engineering, Yale University, 10 Hillhouse Avenue, New Haven, CT 06520
†IBM, Wireless Test Development Group, 1000 River Street 863G, Essex Junction, VT 05452, USA

***Abstract*—In semiconductor manufacturing, a wealth of wafer-level measurements, generally termed inline data, are collected from various on-die and between-die (kerf) test structures and are used to provide characterization engineers with information on the health of the process. While it is generally believed that these measurements also contain valuable information regarding die performances, the vast amount of inline data collected often thwarts efficient and informative correlation with final test outcomes. In this work, we develop a data mining approach to automatically identify and explore correlations between inline measurements and final test outcomes in analog/RF devices. Significantly, we do not depend on statistical methods in isolation, but incorporate domain expert feedback into our algorithm to identify and remove spurious autocorrelations which are frequently present in semiconductor manufacturing data. We demonstrate our method using data from an analog/RF product manufactured in IBM's 90nm low-power process, on which we successfully identify a set of key inline parameters correlating to module final test (MFT) outcomes.**

I. Introduction

During the fabrication process of semiconductor devices, tens of thousands of measurements are collected, from bare silicon all the way through wafer-level to module final test. The data collected during wafer processing is collectively known as *inline data*, and is designed to provide characterization engineers with information about defect density [1], [2], [3] and the electrical/physical properties of the product. This inline data is collected from a variety of test structures. *On-die* structures are located in close proximity to (and may even interact with) the product, with the objective of accurately reflecting its electrical properties, but also with the limitation that available area for such test structures is extremely constrained. *Kerf* structures are located in the wafer kerf, i.e. the areas of the wafer that are destroyed during the wafer dicing step. Depending on the specific product, a large number of test structures (NFETs/PFETs, resistors, capacitors, SRAMs, etc.) are placed in the kerf. Both on-die and kerf structures are sequentially measured throughout wafer processing. To reduce measurement time, however, such inline test structures are not exhaustively tested. Instead, each group of measurements is selectively measured across the wafer to provide a representative sample of the wafer inline measurement statistics.

This incredibly rich dataset is typically used by characterization engineers to monitor the process, control process variation, and identify/respond to problematic processing steps, tools, or off-target process parameters. At first glance, it would seem an easy task to track down the root cause of yield degrades or final test parametric variation given the immense amount of data available from inline test structures. However, it is important to consider the distinction between information and knowledge: from inline test structures, a great deal of information is available to us. Yet without the domain expertise of a number of engineers who sift through and interpret correlations with inline parameters for significance, we are constrained in what we can learn and know about the process.

Indeed, the information sparsity of inline data thwarts automated correlation identification, especially in low-volume products. Consider, for example, the case where we are tasked with establishing a causal link between inline measurements and final test outcomes on a low-volume product. In this scenario, we would be provided with inline data from a small number of wafers, say 100, exhibiting either a yield degrade or parametric variation at final test. The number of inline parameters measured during wafer processing would likely be on the order of 10,000. Yet with such large number of possible predictor variables (i.e., 10,000) and small number of observations (i.e., 100), it would be impossible to construct even a simple linear regression model. In other words, one needs to effectively filter the predictor matrix in order to identify any meaningful correlation to final test outcomes.

Furthermore, it intuitively does not make sense to retain the complete inline dataset when correlating to final test outcomes. The inline test set includes many types of parameters (i.e., simple physical/electrical measurements), some of which may have no physical connection to the final test outcome (i.e., gain, noise figure, IIP3) we are considering. For example, the inline measurement of SRAM beta/gamma ratios in a mixed-signal device may appear correlated to final-test gain of an analog/RF amplifier, but we would not expect a degraded SRAM measurement to cause a degraded gain figure. Thus, implementing a filtering stage is necessary to make the correlation-mining problem tractable.

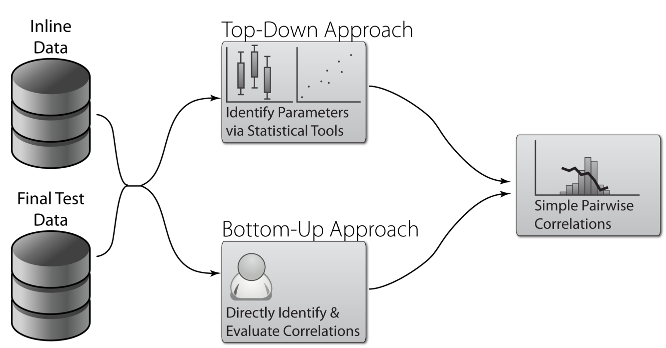


Fig. 1. Current Practice For Identifying Correlations to Inline Parameters

978-3-9810801-7-9/DATE11/©2011 EDAA

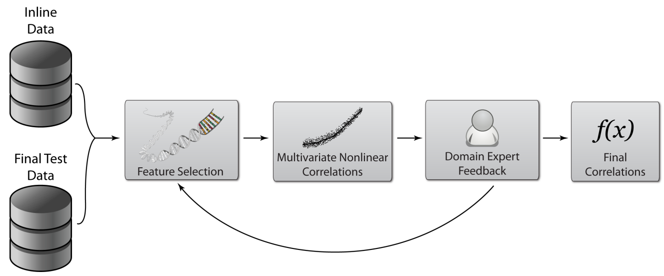


Fig. 2. Overview of Proposed Approach

Characterization engineers traditionally take one of two approaches to the problem of sifting through inline data and tracking down correlations between the inline parameters and final test outcomes, as shown in Figure 1. *Bottom-up*, where domain expertise is employed to manually generate lists of inline parameters and evaluate their correlation to final test, and *top-down*, where statistical tools are employed to automatically identify inline parameters that appear correlated to final test. Each of these approaches has its limitations.

When attempting to identify inline parameters directly via the bottom-up approach, subtle correlations which are not immediately obvious to the domain expert may be overlooked, especially since these correlations may very well be complex functions of multiple inline parameters. The extreme complexity of semiconductor manufacturing means that complete *a priori* knowledge of all relationships between inline parameters and final test outcomes is nearly impossible to achieve.

As for the top-down approach, success is limited by the tools employed: to date, industrial tools for identifying correlation between final test outcomes and inline parameters are univariate or low-dimensionality multivariate and parametric [4]. The limited dimensionality of statistical tools currently deployed in industry limits the correlations that will be uncovered[1]. Moreover, use of parametric statistics relies on assumptions that may or may not be true for the given population. Most significantly, such statistical methods are very sensitive to *spurious correlations*. For example, commonly used statistical tools such as Analysis of Variance (ANOVA) may identify a processing step as affecting final test yield with high statistical significance, but if the suspect step is simply an inspection, say through a Scanning Electron Microscope (SEM), correlation to a yield outcome simply does not make sense.

In this work, we develop a synergistic inline-to-final-test correlation methodology which aims to leverage the strengths of both bottom-up and top-down approaches, while overcoming their limitations. An overview of our proposed methodology is shown in Figure 2. We employ a statistical feature selection approach to consider all inline parameters and combinations thereof, and thereby avoid the limited domain expertise problem which the bottom-up approach is sensitive to. However, we also incorporate engineer domain expertise into our approach, conceding that statistical methods should augment and improve the efficiency of expert intuition, not replace it. Consider, for example, the $100 \times 10,000$ dataset mentioned earlier. With only 100 observations, the probability of finding spurious correlations amongst the 10,000 inline parameters is quite high. By incorporating engineer feedback into our algorithm, we can learn from the domain experts and sidestep these autocorrelations in our analysis.

[1]See [5] for a discussion of cases where presumably redundant parameters provide information gain.

Ultimately, our approach solves two related problems, as shown in Figure 3. First, we dramatically improve the ability of characterization and test engineers to quickly sort through a great number of inline parameters, in order to identify key subsets to monitor when tracking down the root cause of yield degrades or final test parametric variation. Second, once correlations between inline parameters and final test outcomes are well-established via our method, we drastically improve the ability to forecast final test outcomes and, accordingly, plan and fine-tune production schedules. To the best of our knowledge, this work is the first to present a systematic approach to predicting final test outcomes from inline test data, and the first to incorporate feedback from domain experts directly into the algorithm in order to overcome the limitations of the currently practiced top-down and bottom-up approaches.

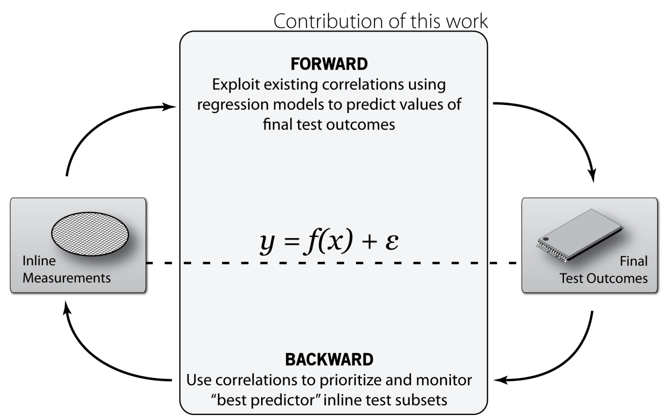


Fig. 3. Utility of Proposed Approach

II. Proposed Correlation Mining Approach

Recall that our objective is to sift through a very large set of inline measurements and identify a small subset which demonstrate good predictive capability in the final test set, i.e., for which we can identify accurate correlation functions with final test outcomes. In addition, we also want to enable domain experts to evaluate the identified correlations and dismiss or reinforce them, with the ultimate goal of enhancing their significance. To this end, three key components are necessary, as shown in Figure 2: (i) a feature selection algorithm, whereby subsets of inline measurements are selected for assessing their effectiveness as predictors of final test outcomes, (ii) a correlation model construction method, whereby the dependent variables (i.e., the final test outcomes) are expressed as functions of the selected predictors (i.e., the inline parameters), and (iii) a provision for domain experts to seamlessly provide feedback and guide an iterative approach, culminating with a final set of correlations which leverage both advanced statistical methods and field knowledge.

A. Feature Selection

Selecting among the large number of inline measurements a subset that "best" correlates to a set of final test outcomes is, essentially, a feature-selection problem. Generally, feature selection is a non-trivial problem, since the number of possible subsets of predictors is $2^p - 1$, where p is the cardinality of the complete predictor set. With even a moderate number of predictors, exhaustive search is completely untenable. Thus, solutions to the feature selection problem generally fall into two classes: greedy methods and heuristic methods. An excellent review of various approaches to feature selection is given in [5]. There is no single "best approach"; each has advantages and disadvantages. Solutions from both classes, have previously been employed by researchers in the analog/RF test community, both in the context of analog/RF specification test compaction [6], [7], [8] and in the context of alternate or machine learning-based test [9], [10]. In our work, we have found heuristic search methods generally work well for the class and size of feature selection problems we encounter; however, one could apply the algorithm presented in this work in conjunction with any other underlying feature selection method with little modification.

The specific heuristic feature selection method employed in this work is a multi-objective genetic algorithm called NSGA-II [11]. Genetic algorithms (GA), also known as evolutionary algorithms, attempt to emulate biological natural selection by creating seed "populations" of solutions, which subsequently undergo mating and mutation steps. These steps are repeatedly performed in phases known as "generations". The justification for such steps are intuitive: by mating two solutions, we may discover a better solution, and by perturbing our solutions via mutation we help avoid local optima which are suboptimal in a global sense. At each generation, every member of the population is evaluated via fitness/objective functions, and the "elites", or best solutions, of each generation are retained. These elites define a Pareto-optimal set of solutions at the termination of the GA, from which we can select a solution optimal for our specific application.

Within the context of inline-to-final-test correlation, the construction of the feature selection problem is straightforward, as we have both a clear objective and a well-defined search space. Our objective is good prediction quality (measured by some loss function on the constructed correlation model) and our search space comprises all possible subsets of one or more inline measurements. As shown in Figure 4, which depicts the first two components of the proposed method, we use NSGA-II to generate bitstrings which correspond to the matrix of inline parameters (i.e., the search space). Each entry in the bitstring determines whether the corresponding inline measurement will be included in the prediction model. The GA then searches through candidate subsets of inline measurements, driven

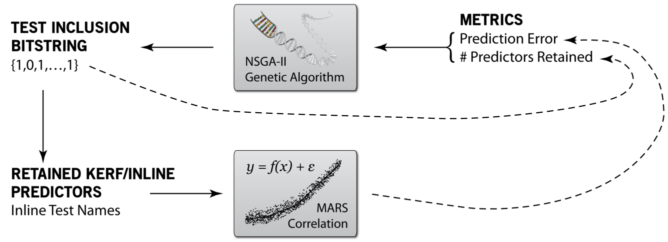


Fig. 4. Iterative Feature Selection and Correlation Model Construction

by two objective functions (recall that NSGA-II is a multi-objective GA). In order to increase information density and avoid the well known "curse of dimensionality" problem, we want to minimize the number of predictors retained, so the first fitness function is proportional to the cardinality of the retained set of inline measurements. We also wish to minimize prediction error, so our second fitness function is proportional to some loss function, which is defined over the final test outcome space. Herein, we use as the loss function the residual sum of squares error of the constructed correlation model.

B. Correlation Model Construction

As shown in Figure 4, at each iteration of the genetic algorithm the selected subset of inline parameters is evaluated by performing a regression to construct a correlation model. In this work, we employ Multivariate Adaptive Regression Splines (MARS) [12] as our correlation model construction method. We use MARS as we have found it consistently outperforms (in terms of prediction error) other regression methods on the semiconductor manufacturing datasets we have been working with. To evaluate prediction error for each subset of inline measurements retained, we build a MARS regression model with our predictor matrix X consisting of the inline measurements, and our dependent variables matrix Y consisting of final test outcomes. Aggregate prediction error is then estimated by the total residual sum of squares (RSS) error across the final test outcomes. To ensure statistical stability, we perform 10 cross-validations and average the RSS. This average is the error which is returned to the genetic algorithm to drive the feature selection process.

C. Domain Expert Feedback

In order to improve convergence time of the genetic algorithm in the large search space of inline measurements, we assign a "prior" over the predictor set based on univariate Pearson correlation coefficients and we modify the algorithm to probabilistically retain parameters based on the prior values. This permits us to loosely prioritize measurements that are well-correlated, in a univariate/parametric sense, to the final test outcomes. At the same time, using this prior does not exclude inline measurements with small correlation coefficients from retention. Thus, we avoid the limitations of simple ranking methods.

Most importantly, this prior also enables us to bridge the gap between top-down and bottom-up methods in a key

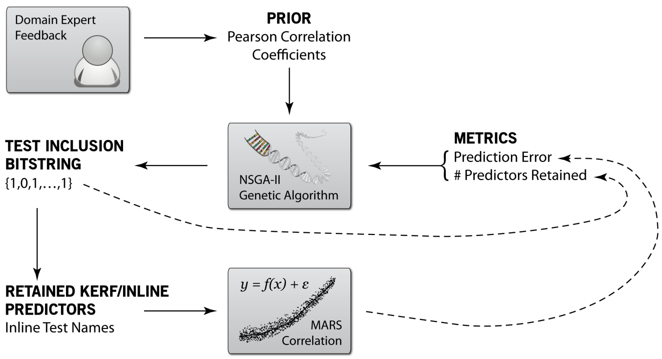


Fig. 5. Complete Correlation Mining Approach

contribution of our method: we can use it to seamlessly incorporate feedback from engineers familiar with the process and continually improve the performance of the feature selection. To do this, we provide simple feedback mechanisms, whereby domain experts can quickly scan the inline measurements retained and flag any correlations which do not make sense given the physical parameters of the device under consideration or any correlations which are highly meaningful and expected. The former are then assigned a very low or zero prior value (i.e., they are likely to be excluded as predictors) while the latter are assigned a very high prior value (i.e., they are likely to be retained as predictors) and the analysis is re-run, generating a new subset of inline parameters correlated with final test outcomes. The complete correlation identification algorithm proposed in this work, complementing feature selection and correlation model construction with probability priors, is presented in Figure 5.

D. Feedback Provision Mechanism

There are many possible mechanisms that can be constructed for domain experts to provide feedback in order to drive the feature selection process. For this work, we opted to implement a web-based interface, the architecture of which is presented in Figure 6.

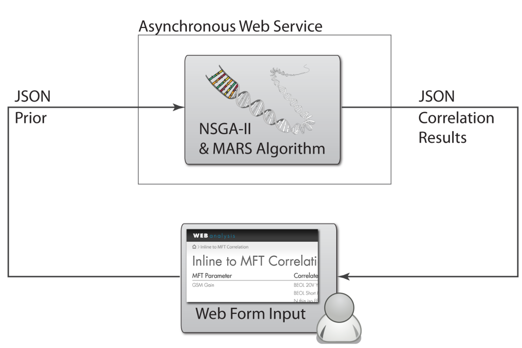


Fig. 6. Domain Expert Feedback Provision Architecture

The algorithm is implemented as an asynchronous web service running on some analysis machine(s). After an initial seed run, results are sent via JSON to a web page. Domain experts can check this page at their convenience for the latest list of correlated inline parameters. For each allegedly correlated parameter, the domain expert can select one of three actions: reject, to remove the inline parameter from the correlation analysis being considered; follow-up, to flag the parameter for further investigation, and accept, to accept the correlation and emphasize its contribution in the prior probability vector. An example of our implementation of this interface is shown in Figure 7.

After feedback is provided by domain experts, the prior is updated asynchronously, again via JSON, and the web service re-runs the analysis, reporting uncovered correlations once again upon completion of the analysis. This can iteratively be repeated with the attention of the characterization engineer only required upon completion of each round of analysis. The anticipation is that, through this iterative process, feedback by the expert engineers will result in more compact and more accurate correlation models.

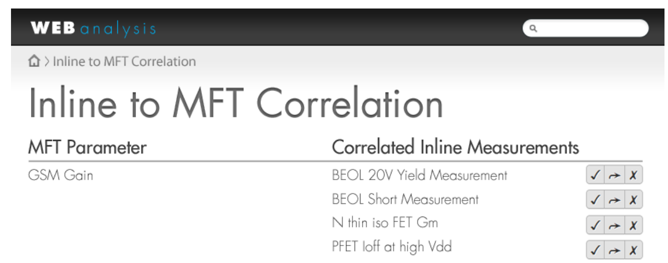


Fig. 7. Example of Web Interface

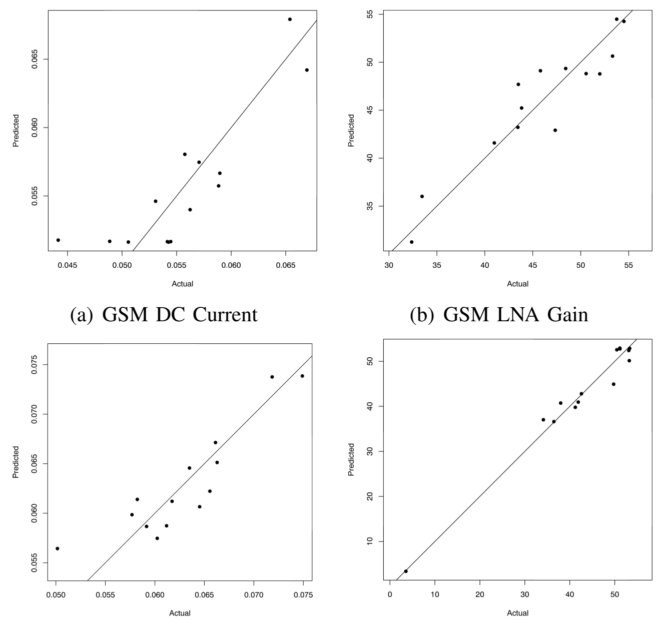
III. Experimental Validation

To evaluate performance of our methodology, we employed a dataset from an analog/RF device manufactured by IBM in their 90nm low-power bulk silicon process, with 14 lots worth of data sampled across several months of production. 14 module final test parameters were identified as key parameters to investigate; these parameters consisted of various supply currents and gains. A large number of inline parameters were also provided; after removing constant columns and columns with missing values, 1,746 inline measurements were retained.

As described in Section I, the contribution of this work can be divided into identification of “forward correlations”, which enable prediction of final test outcomes via subsets of inline tests, and “backward correlations”, which pinpoint key subsets of inline parameters for characterization engineers to investigate and attribute yield degrades and/or final test parametric variations to. Using the IBM-provided dataset, we accomplished both objectives; the results are demonstrated in the following subsections.

A. Forward Correlations: Predicting Final Test Outcomes

Figure 8 provides a graphical presentation of the accuracy of the models constructed by the proposed method for predicting final test outcomes based on inline parameters. Specifically,



(a) GSM DC Current (b) GSM LNA Gain

(c) CDMA DC Current (d) CDMA LNA Gain

Fig. 8. Experimental Results: MFT Parameter Predictions

we show prediction accuracy for 4 out of the 14 final test outcomes, namely a DC current and an LNA gain while operating at the CDMA and the GSM bands. The results for the remaining 12 final test outcomes are similar. Each of the four plots depicts the actual values on the horizontal axis and the values predicted by the constructed models using only inline parameters on the vertical axis. In other words, the 45-degree line shown in the figures represents zero prediction error. As can be seen, we are able to exploit correlations between inline parameters and final test outcomes to successfully predict the latter with minimal prediction error.

The overall prediction error of the proposed method across the 14 parameters, expressed as the residual sum of squares, is provided in Table I and contrasted against two baseline feature selection methods. The first one is a simple rank-based feature selection where the inline measurements are ranked based on their Pearson correlation coefficients, and the most correlated (in a pairwise univariate sense) inline measurements are retained. The second one is a random search, where 1,000 random subsets of inline measurements are evaluated and the retained subset is determined as the subset with the lowest prediction error achieved across all 1,000 iterations. As expected, the heuristic NSGA-II search outperforms both of the simpler feature selection methods, since it efficiently searches the space of inline parameter subsets and uncovers complex multi-variate correlations.

Furthermore, as discussed in Section II, the proposed approach provides the ability for semiconductor manufacturing domain experts to exert fine-grained control over the feature selection process via modification of the prior probability vector, thereby pruning spurious correlations and resulting in compact prediction models using inline parameter subsets of small cardinality. This is corroborated in Table II, which lists

Method	Residual Sum of Squares
Pearson Correlation Coefficient Ranking	1716.980
Random Search - 1,000 Iterations	1243.942
Proposed Feature Selection Method	920.03

TABLE I
EXPERIMENTAL RESULTS: RESIDUAL SUM OF SQUARE ERROR

MFT Parameter	Number of Inline Predictors Retained
DC Current 1	2
DC Current 2	3
DC Current GSM	3
DC Current CDMA	3
Gain GSM	4
Gain CDMA	10

TABLE II
NUMBER OF INLINE PREDICTORS RETAINED

the cardinality of the inline parameter subsets used to predict each of 6 out of the 14 final test outcomes in our experiment.

B. Backward Correlations: Finding Causal Inline Subsets

The key contribution of this work is not in simply discovering correlations, but also in enabling domain experts to identify causal links between inline parameters and final test outcomes. The small number of parameters retained in the correlation models, as shown in Table II, enables quick and effective investigation into the possible root causes of yield degrades and final test parametric variations. In order to achieve compactness of these models, we presented the results of the feature selection and correlation model construction algorithm of Figure 4 to the IBM inline experts and solicited their feedback through the system described in Figure 6.

In Table III, we show the feedback provided through this interaction with the domain experts for 4 out of the 14 final test outcomes. The module final test outcome is listed on the left, the retained inline test parameters correlated to it via NSGA-II and MARS are listed in the center column, and the feedback given by IBM inline experts is provided in the third column. The tendency of automated correlation identification methods to uncover spurious correlations is immediately obvious from the table. Clearly, as we explained earlier, there is no legitimate causal link between SRAM measurements and GSM/CDMA LNA gain, and these correlations are rejected as autocorrelation. On the other hand, the correlation of gain to to NFET transconductance (Gm) makes sense and we expect a legitimate correlation to exist, so we accept such correlations as valid. Other parameters identified as “Follow-up” are parameters flagged for further investigation (i.e., via comparison to broader datasets or other products). Given the collected feedback of the domain experts, the probability priors are updated and the analysis is repeated, with this iterative process ensuring both the accuracy and the compactness of the identified correlation models and causal inline subsets.

IV. CONCLUSION

We presented a novel technique for identifying correlations between inline measurements and final test outcomes, which

Final Test Parameter	Correlated Inline Measurements	Expert Feedback
GSM DC Current	BEOL 10V Yield Measurement	✓Approve
	BEOL 20V Yield Measurement	✓Approve
	N OP RX resistance short device: unsilicided N diffusion resistance	☇ Follow-up
CDMA DC Current	BEOL 10V Yield Measurement	✓Approve
	BEOL 20V Yield Measurement	✓Approve
	Low Vt NFET Idsat	✓Approve
GSM LNA Gain	BEOL Short Measurement	✓Approve
	N thin iso FET Gm	✓Approve
	PFET Ioff at high Vdd	☇ Follow-up
	SRAM PG Vt delta	✗ Reject
CDMA LNA Gain	BEOL Short Measurement	✓Approve
	BEOL 20V Yield Measurement	✓Approve
	N thin iso FET Gm	✓Approve
	Thin gate Coverlap leakage - gate to source/drain extension	✓Approve
	PFET Ioff at high Vdd	☇ Follow-up
	Low Vt PFET Ioff at high Vdd	☇ Follow-up
	PN Igon ratio	☇ Follow-up
	P thin Cov structure inv mode conductance	☇ Follow-up
	Thin NFET Tox Acc to PFET Tox Inv Ratio, N Tox accumulation / P Tox inversion	☇ Follow-up
	SRAM PD/PG ratio: Beta	✗ Reject

TABLE III

CORRELATIONS IDENTIFIED BY NSGA-II FEATURE SELECTION APPROACH & FEEDBACK BY DOMAIN EXPERTS

avoids the pitfalls of both of the traditional bottom-up and top-down approaches to uncovering such correlations. Our approach leverages not only advanced feature selection and correlation model construction methods, but also domain expertise which is seamlessly embedded within our algorithm, thereby providing engineers with a powerful new tool for ensuring accuracy and compactness of the identified correlations. This is especially significant given the prevalence of issues with spurious correlations, which limit the effectiveness and utility of automated correlation identification methods currently used in semiconductor manufacturing. We validated our method on production data from an 90nm analog/RF device manufactured by IBM, demonstrating that our approach can successfully identify both forward correlations, which result in accurate prediction of final test outcomes from inline test parameters, and backward correlations, which pinpoint causal sets of inline parameters for test engineers to monitor and investigate in case of yield degrades and final test parametric variations.

ACKNOWLEDGEMENTS

We would like to thank Roger Kuo, Xuefeng Liu, and Robert Zwonik, all of IBM, for their support with this project and providing domain expertise, and Bill Chang of IBM for reviewing an early draft of this work. We would also like to thank Petros Drineas of Rensselaer Polytechnic Institute for his feedback and input regarding the feature selection algorithm.

REFERENCES

[1] B.T. Murphy, "Cost-size optima of monolithic integrated circuits," *Proceedings of the IEEE*, vol. 52, no. 12, pp. 1537–1545, 1964.

[2] C.H. Stapper, "On Murphy's Yield Integral," *IEEE Transactions on Semiconductor Manufacturing*, vol. 4, no. 4, pp. 294–297, 1991.

[3] J.A. Cunningham, "The use and evaluation of yield models in integrated circuit manufacturing," *IEEE Transactions on Semiconductor Manufacturing*, vol. 3, no. 2, pp. 60–71, May 1990.

[4] G. Kong, "Tool commonality analysis for yield enhancement," in *IEEE/SEMI Advanced Semiconductor Manufacturing Conference*, 2002, pp. 202–205.

[5] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *The Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

[6] S. Biswas, P. Li, R. D. Blanton, and L. Pileggi, "Specification test compaction for analog circuits and MEMS," in *Design, Automation and Test in Europe*, 2005, pp. 164–169.

[7] E. Yilmaz and S. Ozev, "Adaptive test elimination for analog/RF circuits," in *Proceedings of the 46th Annual Design Automation Conference*, 2009, pp. 720–725.

[8] H.-G. D. Stratigopoulos, P. Drineas, M. Slamani, and Y. Makris, "RF specification test compaction using learning machines," *IEEE Transactions on Very Large Scale Integration*, vol. 18, no. 6, pp. 1002–1006, 2010.

[9] H.-G. D. Stratigopoulos and Y. Makris, "Non-linear decision boundaries for testing analog circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 24, no. 11, pp. 1760–1773, 2005.

[10] H.-G. D. Stratigopoulos and Y. Makris, "Error moderation in low-cost machine learning-based analog/RF testing," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 27, no. 2, pp. 339–351, 2008.

[11] K. Deb, A. Pratap, A. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[12] J. H. Friedman, "Multivariate adaptive regression splines," *The Annals of Statistics*, vol. 19, no. 1, pp. 1–67, 1991.