Sparse Regression Driven Mixture Important Sampling for Memory Design

Journal:	*IEEE Transactions on Very Large Scale Integration Systems*
Manuscript ID	TVLSI-00115-2017
Manuscript Type:	Regular Paper
Date Submitted by the Author:	26-Feb-2017
Complete List of Authors:	malik, maria; George Mason University, Electrical and Computer Engg Joshi, Rajiv; IBM, T. J. Watson Res. Center, 11-237, T. J. Watson Res. Center Kanj, Rouwaida; IBM/American University of Beirut, ECE Sun, Shupeng; Google Inc, Software Engineer Homayoun, Houman; George Mason University, Electrical and Computer Engineering Li, Tong; IBM, System and Technology Group
Key Words:	Integrated circuit design, sparse regression, Rare events, statistical analysis, memory

Sparse Regression Driven Mixture Important Sampling for Memory Design

Maria Malik[1], Rajiv V. Joshi[2], Rouwaida Kanj[3], Shupeng Sun[4], Houman Homayoun[1], Tong Li[5]
[1]George Mason University, Fairfax, VA, USA, {mmalik9, hhomayou}@ gmu.edu
[2]IBM T. J. Watson Research Center, Yorktown Heights, NY, USA, rvjoshi@us.ibm.com
[3]American University of Beirut, Lebanon, rouwaida.kanj@gmail.com
[4]Carnegie Mellon University, Pittsburgh, PA, USA, shupengs@ece.cmu.edu
[5]STG, IBM, Austin, TX, tongli@us.ibm.com

ABSTRACT
In this paper, we present a sparse regression based yield analysis methodology and apply it to memory designs with novel write assist circuitry. At the core of engine is the mixture importance sampling methodology which consists of a uniform sampling stage and an importance sampling stage. In our approach, sparse regression model is developed using semi-continuous sample points of the first stage obtained via circuit simulation. The model implements mixture importance sampling for fast and accurate statistical analysis of rare event estimation techniques. The proposed methodology alleviates the need for circuit simulations in the importance sampling phase compared to fully circuit-simulation based approaches. The model-based yield analysis results corroborate well with the circuit-simulation based yield tested on a 14nm FinFET SRAM design. The methodology is used to compare multiple novel SRAM designs including selective boost and write-assist designs. The operating V_{min} ranges and trends corroborate with hardware measurements.

Keywords
Integrated circuit design, memory, SRAM, statistical analysis, rare events, design for manufacturing, sparse regression.

1. INTRODUCTION

With technology scaling, process variations pose a serious challenge to the design and analysis of integrated circuit (IC) design [1-4]. IC generally integrate various circuit components and each component needs to be robust to process variation. Memory designs suffer most leaving serious implication on the chip yield especially for low power design posing further challenges for the operation of portable devices. To achieve the high yield, the failure rate of an SRAM bit-cell must be less than $10^{-8} \sim 10^{-6}$ [13]. With strict requirements of less than one part per million fails, statistical yield analysis techniques have been developed to address the problem of rare event estimation with high confidence [5-7]. The authors in [5] propose mixture importance sampling (MixIS) methods. Unlike Monte Carlo, MixIS avoids simulating too many samples in the success region, instead, it is designed to cover more samples in the tail of the performance distribution.

Gate leakage and standby power consumption analysis are critical for the memory blocks. The significant increase in gate leakage and standby power consumption particularly for memory blocks are necessitating the requirement of powerful techniques. The authors in [6] present a statistical blockade method. Statistical blockade is based on Monte Carlo method that generates sample points in the tails of performance distribution and build advanced models for the tail pdfs. The authors in [7] rely on Gibbs sampling for rare event estimation. Such techniques typically are fully circuit-simulations based.

Sparse regression techniques have been developed [8-12] to address modeling of circuit designs in the presence of variability. In this paper, we explore integration of Orthogonal Matching Pursuit (OMP) [9] method along with importance sampling methods for fast and accurate statistical analysis of rare event estimation techniques. We build a spare regression-based model on the uniform samples to express the model as a function of few critical feature vectors. The resultant model can accurately and efficiently predict the failure points with significant improvement in performance.

An important contribution of this paper is to bypass hundreds to thousands of circuit simulations required in the standard statistical analysis (SSA) technique of SRAM designs to accelerate statistical analysis methodology without compromising the accuracy. The mentioned benefit is exploited by the proposed Sparse Regression Integrated (SRI) statistical analysis technique. The methodology models semi-continuous data. To improve the predictability of the model, we determine the optimal threshold value to minimize the number of false predicts. Later, we integrate the sparse regression model using an optimal threshold value to predict the failures in the context of mixture importance sampling to calculate the yield analysis. Compared to the two-stage process SSA, SRI methodology has the potential to reduce the circuit simulations approximately by half and completely eliminates the second stage of MixIS. This work represents a first application of such an accelerated fast statistical analysis tool to cell functionality. It is implemented in the context of novel selective boosting with write assist circuitry. Selective boosting is applied to the memory and part of logic virtual supply using "single supply". The write assist technique help address the quantization of FinFETs that would otherwise pose problems due to improper cell beta and gamma ratios. The methodology as well as the algorithms uniquely pin point the advantages of the write-assist circuit technique.

The paper is organized as follows. Section 2 presents proposed methodology that includes the mixture importance sampling and OMP method along the cross validation algorithm. Section 3 presents the memory design. Section 4 presents simulation analysis and results. Finally, conclusions are presented in Section 5.

2. PROPOSED METHODOLOGY

In this section, we present the proposed methodology overview and implementation details. Our goal is to enable an accelerated fast statistical analysis methodology. First, we present a review of mixture importance sampling. This enables mapping importance sampling phases to the sparse regression model training and evaluation phases accordingly.

2.1 Mixture Importance Sampling Review

Importance sampling is a variance reduction procedure. It emphasizes on sampling the region of interest. According to [5], it is essential to estimate low failure probabilities like those that follow the Gaussian distribution X, in order for the accurate estimation of yield. The samples that lie towards the tail of the distribution X, are considered the rare events. p(x) is approximately zero in tail region A for which probability P(X ∈ A) is small. The key idea is to produce more samples from the tail region by over-weights the importance region. Instead of the original PDF p(x), sample the distorted PDF g(x). In this case, the failure probability P_f can be expressed as

$$P_f = \int_{-\infty}^{+\infty} \frac{I(x).p(x)}{g(x)}.g(x)dx \qquad (1)$$

where I(x) is the indicator function.

The proper choice of g(x) is critical. g(x) is proportional to the original PDF p(x) [13]. This shows that all failure region should not be sampled uniformly as its efficiency decreases with the increase in dimensionality.

To address this problem, mixed importance sampling has been evaluated in [5] that generates random variables using mixtures of distributions. The algorithm constructs the uniform distribution first and then identify failure points. Next, algorithm determines the center of gravity (μs) by considering only the failure points that lie on one side of the mean. Later, it constructs the new Gaussian curve g(x) = p (x-μs), and integrate the importance sampling. Finally, calculate the yield.

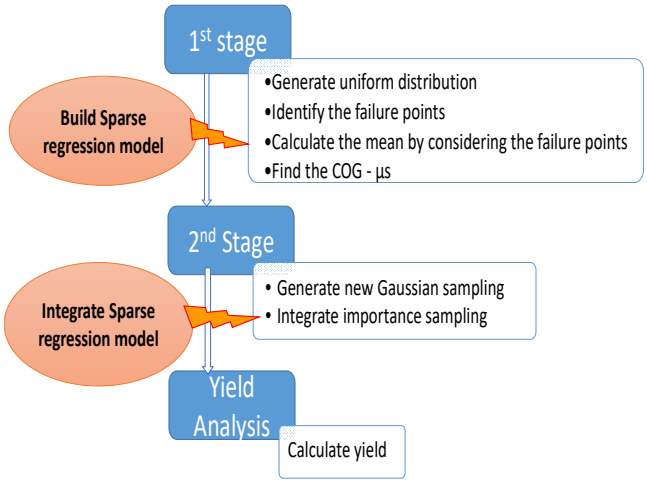


Figure 1. An overview of sparse regression integrated importance sampling.

2.2 Methodology in a Nutshell

Figure 1 presents an overview of our proposed methodology. It employs a mixed simulation/model-based statistical engine, and involves two major steps.

- Use initial set of uniform sample points to build a Sparse Regression model and to calculate the center of gravity (COG) of failure points (μ_s).
- Integrate the Sparse Regression model to predict the failures of the second stage of the mixture importance sample points set.

2.3 Methodology Flow

The methodology flow diagram is presented in Figure 2.

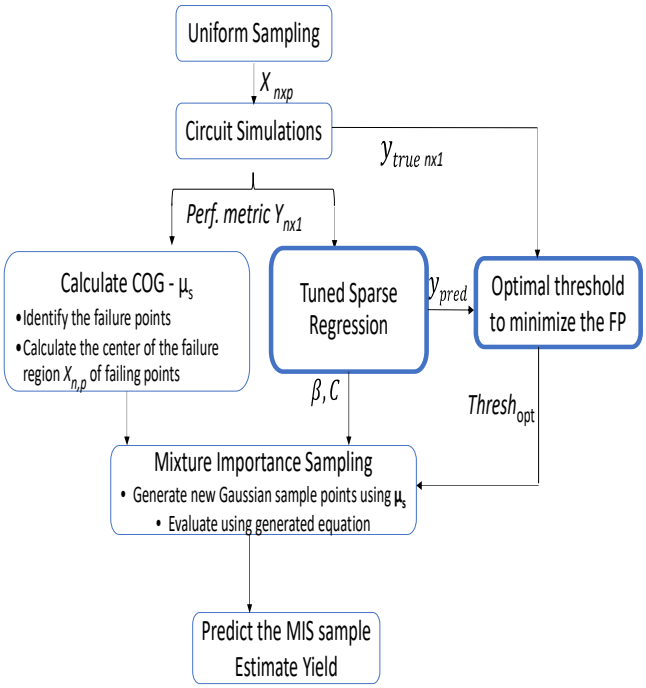


Figure 2. SRI Methodology flow diagram.

Given p explanatory variables or features X_1 …. X_p, and a response variable y. The methodology can be summarized as follows:

Step1: Generate and simulate n sample points uniformly over the p explanatory variables space.

Step 2: Find the center of gravity (COG), μ_s, and use it to shift the natural distribution of the explanatory variables to guarantee that the next stage importance sample points have good coverage of the region of fails.

Step 3: Perform tuned sparse regression for the uniform sample points using OMP and k-fold cross validation technique as illustrated in Figure 3.

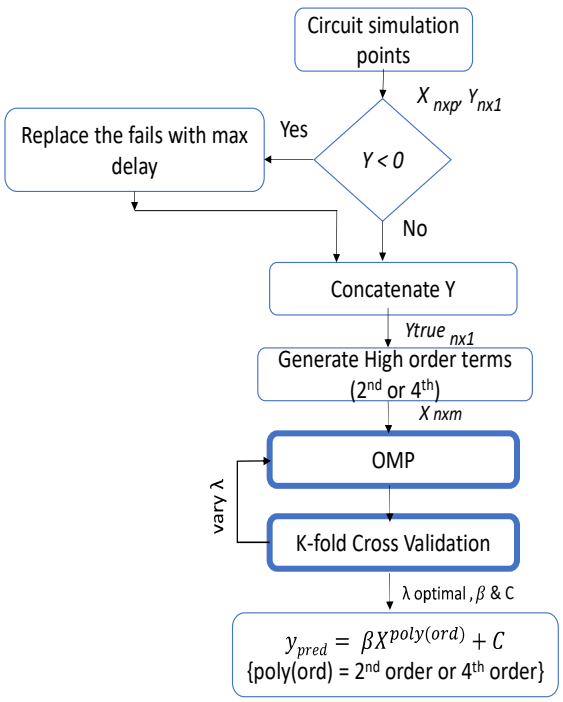


Figure 3: Tuned Sparse Regression flow diagram. High order polynomials include interaction terms.

Given: X: nxp dimensional explanatory variable matrix

Output: Y: nx1 dimensional response variable vector
β: mx1 dimensional model coefficient vector
C: a model scalar entity

Step 3.a: Filter the original response variable (Ytrue) and replace the fail sample points with twice the maximum delay. This makes the data semi-continuous in nature such that data is continuous in the passing region and lumped at the value twice the maximum delay in the failing region. The main reason is this provides separation between the pass and fail regions and allows room for error due to polynomial model fluctuations. Thereby, this enables safety guard for a threshold to separate between the pass and fail regions in the approximate continuous model.

Step 3.b: Generate the higher order polynomial terms for increased model accuracy to enable the model to approximate the complex nonlinear relationship between explanatory and response variables. Thus, this step generates the *nxm* dimensional feature matrix X including interaction variables of the form:

$$X = \{X_i, X_iX_j, X_iX_jX_k, X_iX_jX_kX_l\}$$

$$where\ \{i,j,k,l = 1, \ldots . p\} \qquad (2)$$

Linear regression of the feature matrix with the higher order polynomial terms fits the model with the following equation:

$$Ypred = \beta_1 X_i + \beta_2 X_iX_j + \beta_3 X_iX_jX_k + \beta_4 X_iX_jX_kX_l \ldots + C$$

In general, $Ypred = \beta X^{poly(ord)} + C$

where {poly(ord) = 2nd or 4th order,
X is the extended feature matrix} (3)

Step 3.c: Run the Orthogonal Matching Pursuit (OMP) [9] based regression as explained along with cross-validation.

OMP is used to solve the regularization problem, and accordingly identify the set of critical model features. Thereby producing a high-fidelity sparse model with a few important non-zero coefficients β derived according to (4).

$$\min_{\beta} \left|\left|X\beta + C - Y\right|\right|_2^2 \text{ such that } ||\beta||_0 < \lambda \qquad (4)$$

where the performance Model is:

$$y = f(x) = \sum_{i=1}^{m} \beta_i * X_i + C$$

λ is the regularization parameter that controls the sparsity level using cross-validation technique. The general modifications to the equation (4) are L0-norm, L1-norm and L2-norm regularization [10].

OMP can be summarized by the following steps:

1. Initialize the residual to be R=y; and the set of selected features to be Ω={}
2. Select from '{X}-Ω' the feature with the highest correlation to R.
3. Build regression model for y using only the selected features in Ω.
4. Update R based on the new model. If card (Ω) did not exceed λ, go to step 2.

As mentioned earlier, regularization results depend on λ, which can be tuned automatically based on the model error as determined by cross-validation. We have considered 5-fold cross-validation, in our proposed model. In 5-fold cross-validation, the input dataset is split into 5 mutually exclusive folds. A single fold is considered as the test set while remaining 4 folds are held out for training sets. The cross-validation process runs 5 times. To consider all the input data for both testing and training, each of the 5 folds is evaluated at one instance as a test data. Figure 4 illustrates the case for 5-fold cross-validation.

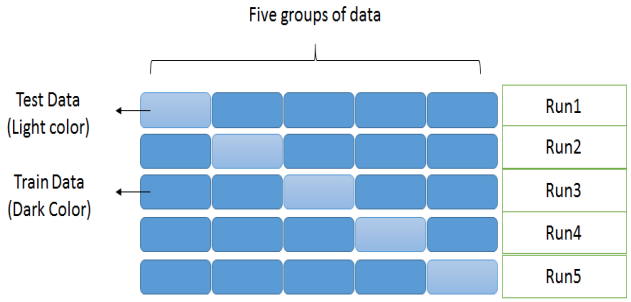


Figure 4. 5-fold cross validation.

Repeat step 3.c for increasing values of λ until we find the optimal λ (number of sufficient significant model coefficients). We stop the search once λ exceeds a certain

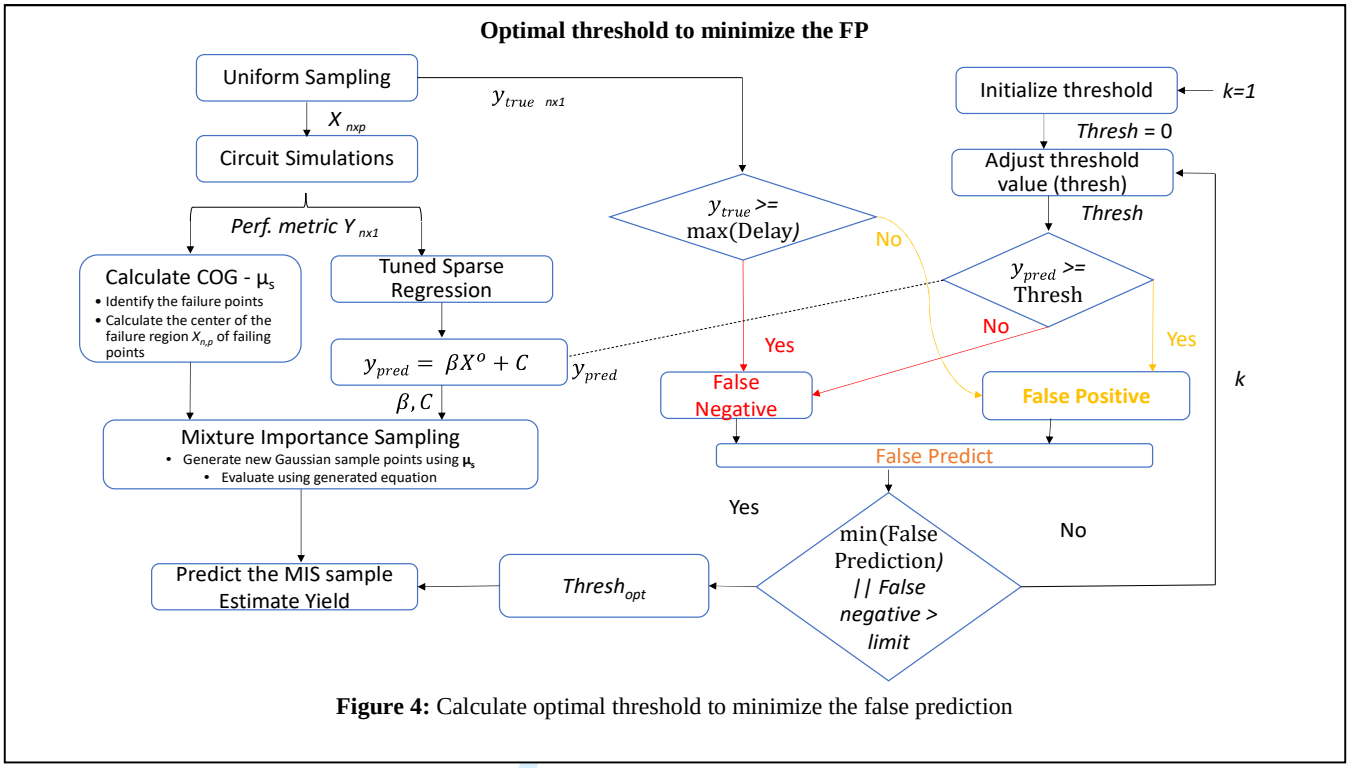


Figure 4: Calculate optimal threshold to minimize the false prediction

maximum limit predefined by the user (for example 30 features at max).

Step 3.d: Predict the response variables (Ypred) using the optimal coefficient vector β and scalar entity C as calculated in Step 3.c

Step 4: To improve the predictability of the model, we tune the fail threshold to minimize the overall number of false predictions while limiting the false negatives to a certain value. We refer to this fail as the optimal threshold value that properly maps the uniform sample points true pass/fails to the predicted pass/fail. The optimal threshold is then used for identification of pass/fail values for the predicted importance sample points.

Step 4.a: Initialize the threshold value (Thresh)

Threshold value is considered as a limit for the pass and fail criteria. Any predicted sample points (Ypred) greater than threshold value are considered as fail.

Step4.b: Identify false predict

False predict is the sum of the false positive and false negatives that are calculated using conditions as follows:

False Negative } Ytrue >= max (Delay) and Ypred < Thresh

False Positive } Ytrue < max(Delay) and Ypred >= Thresh

Step 4.c: Search for the optimal threshold value to obtain the minimum number of false predict. The value of the threshold is incremented gradually to identify the minimum number of false predictions within a maximum value of the number of false negatives.

Step 5: Generate importance sample points using the shifted Gaussian (natural) distribution, and evaluate the performance metric using the model coefficients, β and scalar entity C as generated in step 3 and optimal threshold value discussed in Step 4.

Step 6: Predict the failures in MixIS and compute the yield using equations similar to equation (1).

3. Application to Newly Developed Write Assist Circuitry

Non-planar technology (FinFET) brings forth a new challenge in terms of the quantization of FinFETs which poses a problem for proper beta and gamma ratios used in the cell. A minimum sized SRAM cell with 1 fin each for all the devices drives a beta and gamma ratio of 1. This disturbs the "write-ability" at low "Vdd". To overcome such problems new circuit techniques are essential. In this paper, we study unique selective transistor-based boosting circuit techniques along with write assist techniques for the purpose of improving low voltage operation.

Figure 5 presents a typical boost circuit. It involves an nFET in parallel with a pFET device. The pFET source and nFET drain are connected to Vdd. When their common gate is switched from low to high, it couples the nFET's source (pFET's drain) to a value above Vdd. The devices are sized to give a boost around 0.12V at low voltage operation.

The selective boost technique allows the design to operate at lower Vdd values by selectively boosting memory

specific paths and excluding surrounding logic from the boost. The selective boosting can be further paired with other write assist techniques. We revise two such techniques [14].

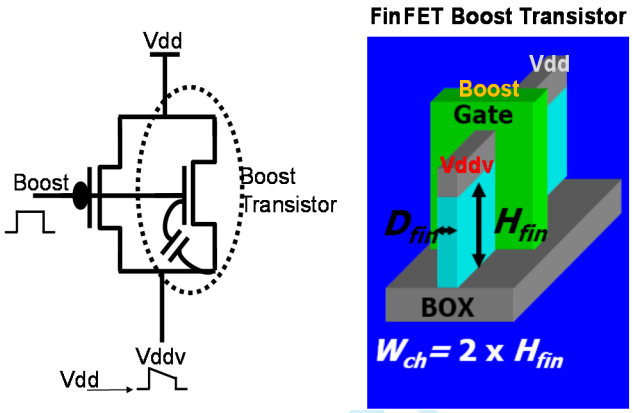


Figure 5. Capacitive Coupling between Gate and source boosts the source voltage (Vddv) above Vdd when Boost Switches "High" [14]

The first technique, described in Figure 6 works with voltage collapse [14]. The boost is on and Vddv is supplied for the wordline path, bitlines as well as the cell. During write, the technique reduces the memory cell supply voltage. Hence, when the write clock is on, the cell voltage drops with one threshold voltage, Vt, lowering the cell voltage and making cell writability more feasible.

The second write assist scheme [14], relies on negative boosting of the bitlines (Figure 7). The negative boost is created through capacitor coupling between the gate nodes of datat (data true) and datac write enable transistors and the bitlines. During the write operation, the negative boost brings the bitline voltage lower than zero and generates an increased voltage swing between the true and complement bitline during write. Thus, the increase at the gate voltage makes the transistor strong so it can flip the cell bit easily and enhance the writeability.

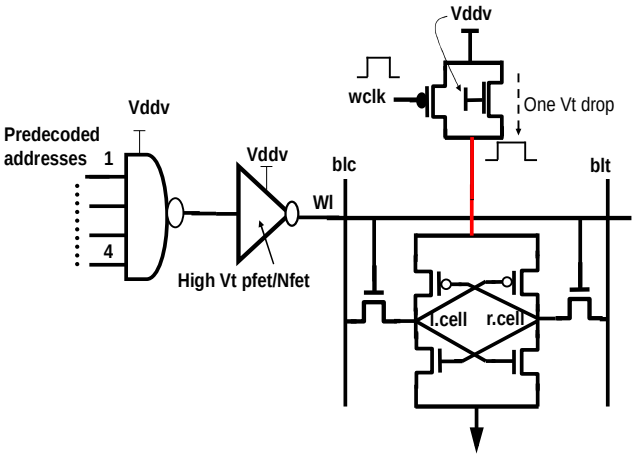


Figure 6. Voltage Collapse Write Assist Technique – Applied to SRAM cell with Virtual supply (Vddv) [14].

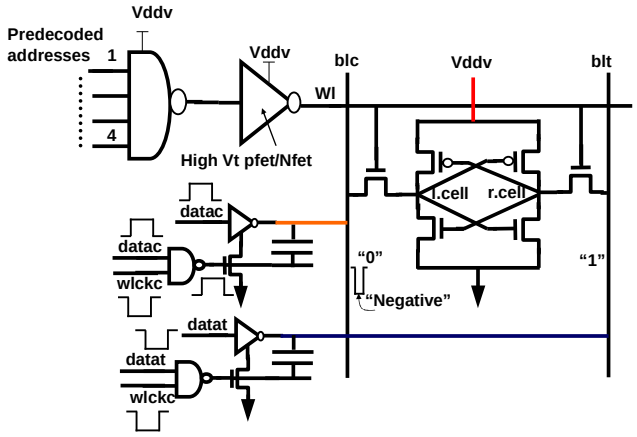


Figure 7. Negative Boost Write Assist - Applied to bitlines with Virtual supply (Vddv) [14]

Selective boost applied to selective paths– wordline, write drivers and cell, therefore, improves the yield and pushes the low voltage operation range. The write assist techniques further help the SRAM cells improve the write-ability yield. Tables I lists three design options that were implemented in [14] for a 14nm FinFET SOI technology 72Kb SRAM array arranged in columns of 16 cells/bitline. The presented techniques are as follows:

1. Selective boosting only
2. Selective boosting paired with voltage collapse
3. Selective boosting paired with negative bitline boost

Table I. Designs understudy

Design	Write Assist	Referred as
Selective Boost Only	None	SB_NoAssist
Selective Boost with Write Assist	Voltage Collapse	SB_Collapse
	Negative Bitline Boost	SB_NegBoost

Hardware measurements indicate that SB_NoAssit design requires more than 0.45V to operate. SB_Collapse allows the cells to work with the low voltage as of 0.35V. SB_NegBoost further stretches the operating voltage range all the way to 0.30V for write-ability. These Vmin operating ranges and trends were designed and validated using statistical design methodologies [5] and results were found to corroborate well with fabricated hardware presented in Figure 8. Figure 9 illustrates the yield improvement due to SB_Collapse and SB_NegBoost compared to the SB_NoAssit. It is clear that write assist SB_NegBoost provides significant yield improvement compared to SB_Collapse as it can operate at the lower supply voltage.

4. ANLAYSIS AND RESULTS

For purposes of our analysis, we demonstrate the efficacy of the proposed SRI methodology in the evaluation of the negative bitline boost write assist (SB_NegBoost)

technique in terms of model prediction capability and yield estimation.

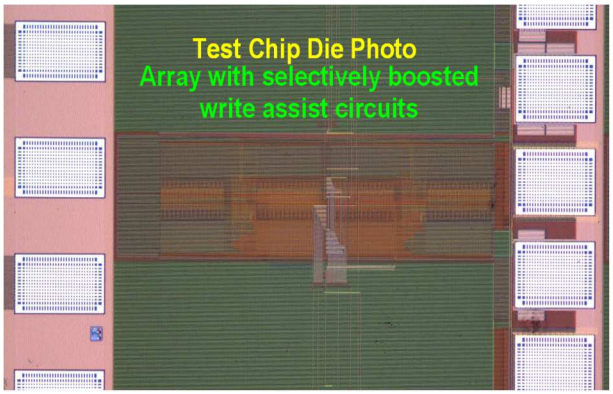


Figure 8. Die photo in 14nm FinFET SOI technology

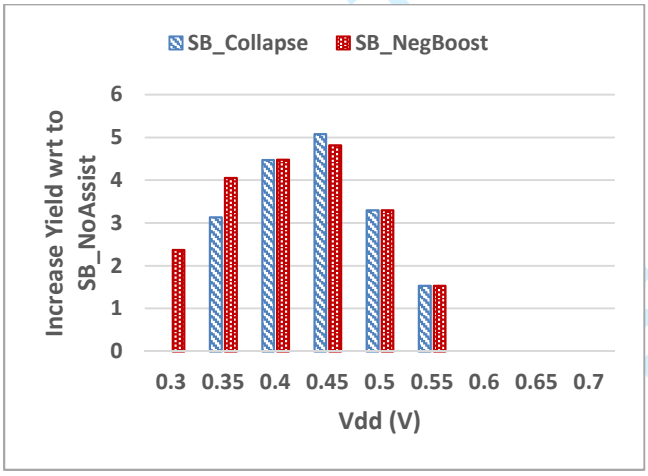


Figure 9. Increase in yield with SB_Collapse and SB_NegBoost compared to the SB_NoAssist (arbitrary units)

4.1 Experimental Setup

We apply the methodology to an industrial 14nm FinFET SRAM design. For accurate analysis, simulations involve the cell along with the peripheral logic as illustrated in Figure 10. Variability is injected in the memory cell devices as well as the local evaluation circuitry. Variability effects such as metal gate granularity, line edge roughness, fin height variation and random dopant fluctuations are lumped into one source σ_{vt} that is injected into the simulations. Write-ability, which is the ability of the cell to be written, is selected as the primary metric for investigation.

4.1 Model Building

We have generated 1000 samples using uniform distribution for 9 studied features/explanatory variables. Due to the nonlinearity of memory designs, we build a higher order polynomial model (2nd order or 4th order) to improve the accuracy. To handle the sparsity, OMP expresses the model as the function of few important non-zero coefficients β. We employ a 5-fold cross-validation approach for optimal λ. For instance, with the 4th order polynomial model OMP focuses only on 30 non-zero coefficients β as compared to a full blown model with 714 terms.

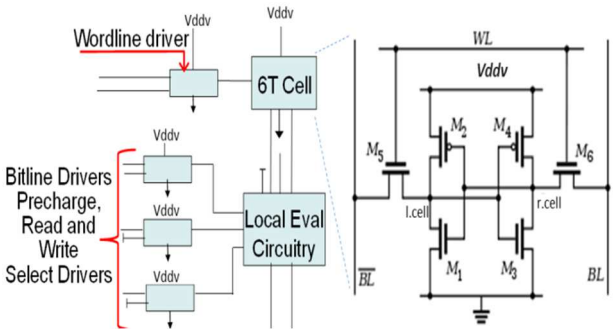


Figure 10. SRAM cell cross-section. 6T portion of 8T cell illustrated

4.2 Optimal Threshold Selection Analysis

We determine the threshold value used to predict the pass and fail criteria for the 2nd MixIS stage as the value that minimizes the 1st stage false predictions (see Figure 2 and Figure 4). Figure 11 presents an example optimal threshold value, Threshopt. Threshold range can be divided into two regions; overpredict cell fails – when threshold values are lower than the Threshopt - and underpredict fails - when threshold values are over the Threshopt value. In fact, for values smaller than Threshopt the model is pessimistic and has a large number of False Positives (False Fails), whereas for values larger than Threshopt, the model is optimistic and the number of False Negatives (False Passes) is high. False predict is the summation of false positives and false negatives. We select the optimal threshold value that provides the minimum false predict and limits the false negatives within a certain range. In this paper, for a conservative model, we have considered less than or equal to 10 false negatives.

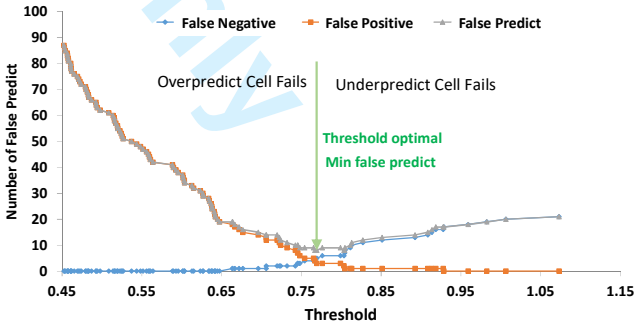


Figure 11: Optimal Threshold Selection. Threshold < threshold optimal: overpredicting cell fails. Threshold > threshold optimal: underpredicting cell fails.

4.3 MixIS Model Prediction Analysis

We apply the proposed methodology to analyze the designs understudy as the supply voltage is varied over the range [0.40V – 0.43V] for the SB_NoAssist design and [0.35V-0.40V] for the SB_NegBoost design. In this section,

we focus on evaluating the proposed model prediction capabilities for the final MixIS stage. We present the percentage of false positives and false negatives (uncaptured fails) for both designs in Figure 12 and Figure 13 respectively. The results are for 2nd order and 4th order polynomial model.

Overall, we find that the 4th order model has better prediction capability for both designs. We report the results in terms of the percent false positive and percent false negatives compared to the true number of fail points. For the 2nd order polynomial, SB_NoAssist design, we observe that false positives percentage is as small as 0.4% for all the voltages. The maximum error is 20% for the false negative at 0.43V supply voltage. On the other hand, the maximum error for the 4th order polynomial is 6% recorded for the false positives at 0.43V.

For the SB_NegBoost design, again the 4th order polynomial demonstrates lower errors in comparison to the 2nd order. The maximum recorded error is 15% for the false negatives recorded at 0.38V.

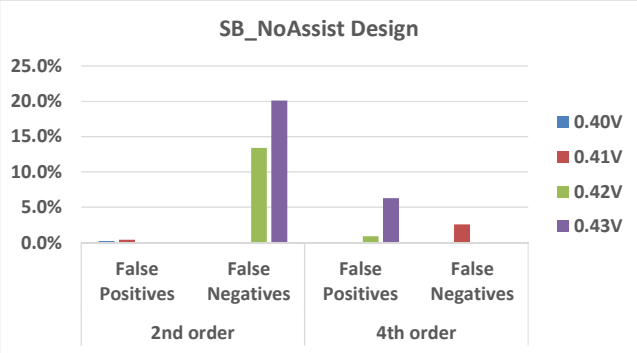


Figure 12: MixIS Prediction Summary for SB_NoAssist Design

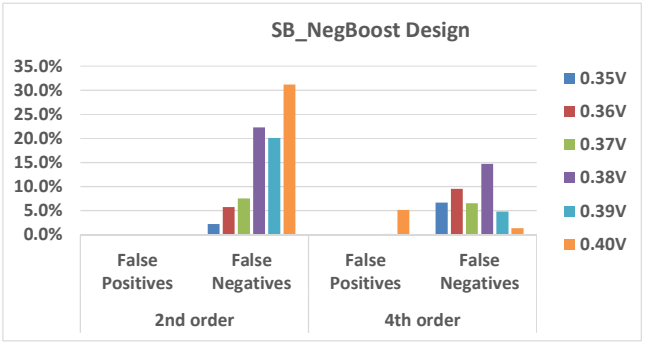
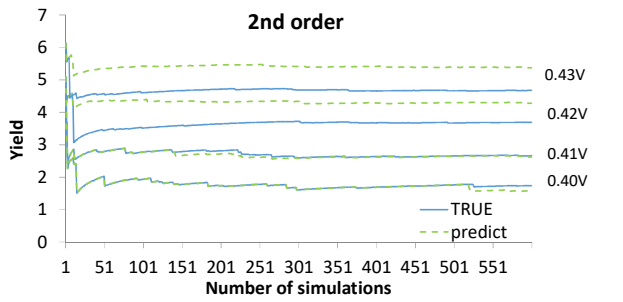


Figure 13: MixIS Prediction Summary for SB_NegBoost Design

4.4 Yield Convergence Analysis

We apply the SRI methodology to estimate the yield of the SB_NoAssist and SB_NegBoost memory design. We analyze the convergence of the estimate with respect to the number of simulations. The yield estimation is based on the 2nd order and 4th order polynomial models presented in the previous section. The respective yield convergence results are presented in Figure 14 and Figure 15. We observe that for both SB_NoAssist design and SB_NegBoost design, 2nd order polynomial does not provide proper yield convergence especially at the high supply voltage. This is expected due to the higher errors observed for the 2nd order model prediction presented in the section 4.3. However, 4th order model yield estimation results demonstrate high corroboration between the proposed model-based SRI methodology and the traditional circuit-simulation based yield technique. This is observed at both low and high supply voltages.

In addition, we present the final model based yield in comparison to the actual yield for the SB_NoAssist and SB_NegBoost designs, as shown in the Figure 16 and Figure 17. It is clear that the 4th order polynomial model estimates



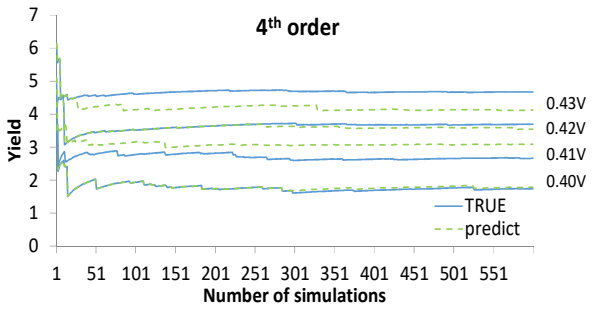
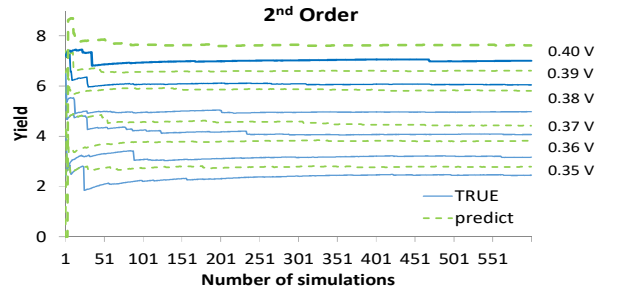


Figure 14: Yield Estimation: SB_NoAssist Design- Convergence



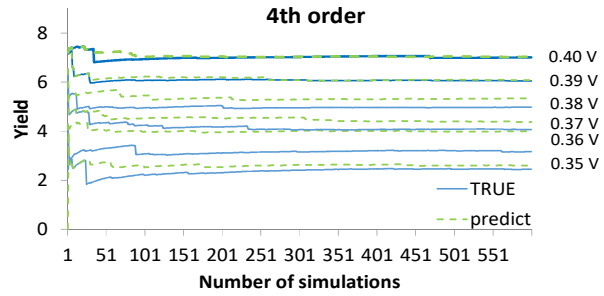


Figure 15: Yield Estimation: SB_NegBoost Design- Convergence

yield more accurately compared to the 2nd order polynomial for both designs.

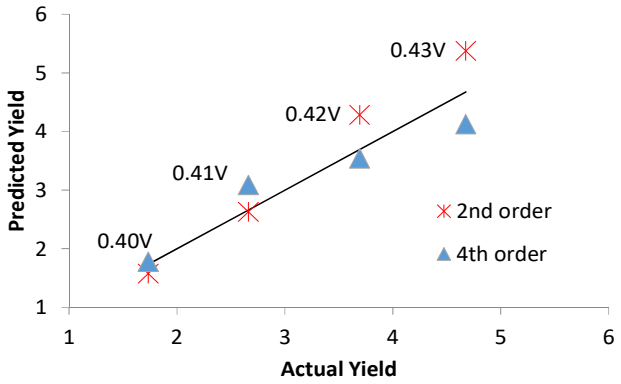


Figure 16: SB_NoAssist Design Model-based Yield Estimation arbitrary (on the axis)

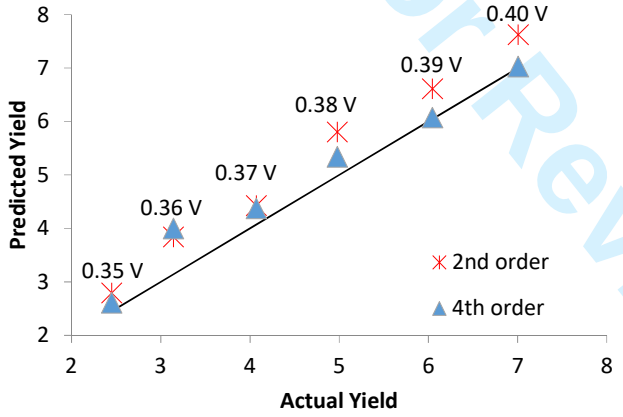


Figure 17: SB_NegBoost Design Model-based Yield Estimation arbitrary (on the axis)

The proposed technique can be effectively applied to other chip design frameworks in general. Considering that SRI reduced the circuit simulation by half and completely eliminated the need of circuit simulation for the second stage of mixture important sampling, it enables significant speedup compared to the traditional circuit-simulation based technique.

5. CONCLUSIONS

We present a first application of sparse regression based statistical methodology for rare fail estimation of memory designs. In the proposed methodology SRI, we integrate the sparse regression model using an optimal threshold value for semi-continuous data to predict the failures in the context of mixture importance sampling for yield analysis calculations. The methodology is shown to efficiently model cell functionality in the context of high dimensional space with application to novel selective boosting with write assist circuitry. SRI bypasses hundreds to thousands of circuit simulations by completely eliminating the second stage of MixIS required in the standard statistical analysis (SSA) technique of SRAM designs to accelerate statistical analysis methodology without compromising the accuracy of the model.

6. REFERENCES

[1] V. De, S. Borkar, "Technology and design challenges for low power and high performance [microprocessors]," in *Proceedings. 1999 International Symposium on Low Power Electronics and Design*, pp.163-168.1999.

[2] S. R. Nassif, "Design for variability in DSM technologies [deep submicron technologies]," in *Proceedings IEEE 2000 First International Symposium on Quality Electronic Design, 2000. ISQED 2000*, pp.451-454. 2000.

[3] R.V. Joshi, Y. Chan, D. Plass, T. Charest, R. Freese, R. Sautter, W. Huott, U. Srinivasan, D. Rodko, P. Patel, P. Shephard, T. Werner, "A Low Power and High Performance SOI SRAM Circuit Design with Improved Cell Stability," in *International SOI Conference, 2006 IEEE* , vol., no., pp.4-7. 2006.

[4] C. Visweswariah, "Plenary Speech 2P2: Statistical Techniques to Achieve Robustness and Quality," in *9th International Symposium on Quality Electronic Design, ISQED 2008*, pp.586-586.2008.

[5] R. Kanj, R. Joshi, S. Nassif, "Mixture importance sampling and its application to the analysis of SRAM designs in the presence of rare failure events," in *43rd ACM/IEEE Design Automation Conference*, pp.69-72. *2006*.

[6] A. Singhee, R.A. Rutenbar, "Statistical Blockade: A Novel Method for Very Fast Monte Carlo Simulation of Rare Circuit Events, and its Application," in *Design, Automation & Test in Europe Conference & Exhibition, DATE '07* , pp.1-6. 2007.

[7] Changdao Dong, Xin Li, "Efficient SRAM failure rate prediction via Gibbs sampling," in *48th ACM/EDAC/IEEE Design Automation Conference (DAC), 2011*, pp.200-205.

[8] X. Li, J. Le and L. Pileggi, "Statistical performance modeling and optimization," *Foundations and Trends in Electronic Design Automation*, pp. 331-480, 2006.

[9] X. Li and H. Liu, "Statistical regression for efficient high-dimensional modeling of analog and mixed-signal performance variations," *DAC*, pp. 38-43, 2008.

[10] X. Li, "Finding deterministic solution from underdetermined equation: large-scale performance modeling by least angle regression," *DAC*, pp. 364-369, 2009.

[11] Y. Wang, M. Orshansky and C. Caramanis, "Enabling efficient analog synthesis by coupling sparse regression and polynomial optimization," *DAC*, 2014.

[12] Y. Zhang, S. Sankaranarayanan and F. Somenzi, "Sparse Statistical Model Inference for Analog Circuits under Process Variations", *ASP-DAC*, pp. 449-454, 2014.

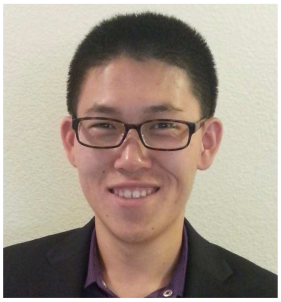
[13] Sun, Shupeng, Xin Li, Hongzhou Liu, Kangsheng Luo, and Ben Gu. "Fast statistical analysis of rare circuit failure events via scaled-sigma sampling for high-dimensional variation space." In Proceedings of the International Conference on Computer-Aided Design, pp. 478-485. IEEE Press, 2013.

[14] R. V. Joshi, M. Ziegler, H. Wetter, C. Wandel, H. Ainspan," 14nm FinFET Based Supply Voltage Boosting Techniques for Extreme Low Vmin Operation," VLSI Circuit Symp., 2015, pp. C268

Maria Malik is currently working towards the Ph.D. degree in Electrical and Computer Engineering department, at George Mason University, VA. She has received the M.S. degree in Computer Engineering from the George Washington University, DC and B.E. degree in Computer Engineering from the Center of Advanced Studies in Engineering, Pakistan. Her research interests are in the field of Computer Architecture with the focus of performance characterization and energy optimization of big data applications on the high performance servers and low-power embedded servers, accelerating machine learning kernels, parallel programming languages and parallel computing.

Rajiv V. Joshi is a research staff member at T. J. Watson research center, IBM. He received his B.Tech I.I.T (Bombay, India), M.S (M.I.T) and Dr. Eng. Sc. (Columbia University). He holds 58 invention plateaus and has over 225 US patents and over 350 including international patents. He has authored and co-authored over 185 papers. He received the Best Editor Award from IEEE TVLSI journal. He is recipient of 2015 BMM award. He is inducted into New Jersey Inventor Hall of Fame in Aug 2014 along with pioneer Nikola Tesla. He is a recipient of 2013 IEEE CAS Industrial Pioneer award and 2013 Mehboob Khan Award from Semiconductor Research Corporation. He is a member of IBM Academy of technology. He served as a Distinguished Lecturer for IEEE CAS and EDS society. He is IEEE, ISQED and World Technology Network fellow and distinguished alumnus of IIT Bombay. He is in the Board of Governors for IEEE CAS. He serves as an Associate Editor of TVLSI. He served on committees of ISLPED (Int. Symposium Low Power Electronic Design), IEEE VLSI design, IEEE CICC, IEEE Int. SOI conference, ISQED and Advanced Metallization Program committees.

Rouwaida Kanj received the M.S. and Ph.D. degrees in electrical engineering from the University of Illinois Urbana-Champaign in 2000 and 2004, respectively. She is currently an Assistant Professor at the American University of Beirut. From 2004-2012 she worked as a research staff member at IBM Austin Research Labs. She worked on modeling SOI effects, noise characterization of CMOS circuits, library characterization of novel circuit technologies, and is currently involved in statistical design methodologies. Dr. Kanj was a recipient of three IBM Ph.D. Fellowships, is the author of more than 55 technical papers, 20 issued patents and several pending patents. She received an outstanding technical achievement award and 6 Invention Plateau awards from IBM. She received the IEEE/ACM WILLIAM J. MCCALLA ICCAD best paper award in 2009, and ISQED best paper award in 2006 and 2014, and is currently a senior member of IEEE.award in 2009, and IEEE ISQED best paper award in 2006 and 2014, and is currently a senior member of IEEE.



Shupeng Sun received the B.E. degree from the Department of Automation, Tsinghua University, Beijing, China in 2010, and the Ph.D. degree from the Department of Electrical and Computer Engineering, Carnegie Mellon University, PA, USA in 2015. He is currently a Software Engineer at Google Corporation, Mountain View, CA, USA. Dr. Sun was a recipient of the IEEE Donald O. Pederson Best Paper Award in 2016, and the ACM SIGDA DAC Ph.D. forum Best Poster Research Award in 2014. He got the Gold medal in ACM student research competition at ICCAD in 2014, and the second place in ACM student research competition grand finals in 2015.

Houman Homayoun is an Assistant Professor of the ECE department at George Mason University. He also holds a joint appointment with the Computer Science as well as Information Science and Technology departments. He is the director of GMU's Green Computing and Heterogeneous Architectures (GOAL) Laboratory. Prior to joining GMU, he spent two years at the UC San Diego, as National Science Foundation Computing Innovation (CI) Fellow awarded by the CRA and CCC. Houman is currently leading a number of research projects, including the design of next generation heterogeneous multicore accelerator for big data processing, non-volatile STT logic, heterogeneous accelerator platforms for wearable biomedical computing, and logical vanishable design to enhance hardware security which are all funded by National Science Foundation (NSF), General Motors Company (GM) and Defense Advanced Research Projects Agency (DARPA). Houman received his PhD degree from the Department of Computer Science at the University of California, Irvine in 2010, an MS degree in computer engineering in 2005 from University of Victoria, Canada and his BS degree in electrical engineering in 2003 from Sharif University of technology.

Tong Li, graduated from EE dept of Xian JiaoTong University in 1989 and 1992 with Bachelor and Master degree, and EE dept of New Jersey Institute of Technology in 1999 with Ph.D degree. Since joined IBM EDA in 2004, have been worked in the area of circuit simulation, fast transistor model generation, and yield analysis of circuit design.