REscope: High-dimensional Statistical Circuit Simulation towards Full Failure Region Coverage

[1]Wei Wu, [2]Wenyao Xu, [1]Rahul Krishnan, [1,3]Yen-Lung Chen, [1]Lei He
[1]EE Dept., University of California, Los Angeles, CA, USA
[2]CSE Dept., University of Buffalo, SUNY, NY, USA
[3]EE Dept., National Central University, Taiwan, R.O.C
weiw@seas.ucla.edu, wenyaoxu@buffalo.edu, r.krishnan390@ucla.edu, 995401004@cc.ncu.edu.tw,
lhe@ee.ucla.edu

ABSTRACT

Statistical circuit simulation is exhibiting increasing importance for circuit design under process variations. Existing approaches cannot efficiently analyze the failure probability for circuits with a large number of variation, nor handle problems with multiple disjoint failure regions. The proposed rare event microscope (REscope) first reduces the problem dimension by pruning the parameters with little contribution to circuit failure. Furthermore, we applied a nonlinear classifier which is capable of identifying multiple disjoint failure regions. In REscope, only likely-to-fail samples are simulated then matched to a generalized pareto distribution. On a 108-dimension charge pump circuit in PLL design, REscope outperforms the importance sampling and achieves more than 2 orders of magnitude speedup compared to Monte Carlo. Moreover, it accurately estimates failure rate, while the importance sampling totally fails because failure regions are not correctly captured.

Categories and Subject Descriptors: B.7.2[Integrated Circuits]: Design Aids – Simulation

General Terms: Algorithms, Design

Keywords: Circuit simulation, Yield Estimation, Process variation, Monte Carlo methods, Classification

1. INTRODUCTION

As electronic devices scale to much smaller sizes than ever before, circuit reliability has become an area of growing concern due to the uncertainty during IC manufacturing. For critical circuits, such as PLLs, which stabilize the clock for the entire chip, and RAM cell, which is duplicated for millions of times, an extremely small failure probability may cause a catastrophe for the entire chip. Traditional circuit simulation performs deterministic worse case analysis (WCA) to decide a safety margin during the design. It is, however, not sufficient to analyze the rare failure event [1].

Modern circuit simulation takes process variations into

account and statistically simulates the probability that a circuit does not meet the performance metric. A "gold standard" approach to reliably estimate the probabilistic circuit performance is Monte Carlo (MC) [2], which repeatedly draws samples and evaluates circuit performance with transistor-level SPICE simulation. A lot of efforts have been spend to reduce the runtime of a single simulation [3, 4, 5]. However, MC is still extremely inefficient because millions of samples need to be simulated to capture one single failure when the failure is a rare event.

To mitigate the inefficiency issue of MC method, fast statistical approaches have been proposed in the past decade, which can be categorized in the following groups:

(1) **Moment matching** [6, 7, 8]: The approaches in this category only evaluate a small number of samples with SPICE simulation, and approximate the PDF of the circuit performance to an analytical expression by means of moment matching. However, existing moment matching based approaches are known as numerically instable because the moment matrix solved during moment matching is usually ill-conditioned [9, 10]. Moreover, they only match the overall shape of the PDF without surgically looking into its tail, which contains information special to rare events. Therefore, these algorithms are usually applied to low dimensional behavior modeling rather than high-dimensional rare event analysis.

(2) **Importance Sampling**: To specifically look into the samples that cause a rare event, importance sampling based approaches [11, 12, 13, 14, 15] have been developed to construct a new "proposed" sampling distribution under which a "rare event" becomes less rare so that more failures can be easily captured. The critical issue is how to build an optimal "proposed" sampling distribution. For example, [11] mixes a uniform distribution, the original sampling distribution and a "shifted" distribution centering around the failure region. The approaches in [12, 13, 15] shift the sampling distribution towards the failure region with a minimum L_2-norm. The work in [14] uses "particle filtering" to tilt more samples towards the failure region. However, all these approaches are related to mean shifting and they are based on the assumption that all failed samples are located in a single failure region. In reality, the "proposed" sampling distribution may not effectively cover all the failed samples when they are spread in multiple disjoint failure regions.

(3) **Classification**: the approach in statistical blockcade (SB) [1, 16] makes use of a "classifier" to "block" those Monte Carlo samples that are unlikely to cause failures and simu-

lates the remaining samples. However, the linear support vector machine (SVM) used in SB can be easily fooled in high dimensional [15], nor can it effectively deal with multiple failure regions. In particular, if the existence of multiple failure regions is known, it has to use multiple classifiers to find out each failure region respectively [17]. However, the existence of multiple failure regions is blind to the algorithm under most cases.

Clearly, most of the existing approaches can be successfully applied to low-dimensional problems with small number of variables, but, in general, perform poorly in high dimension. Moreover, none of these approaches considers efficiently handling the samples in multiple disjiont failure regions.

The proposed rare event microscope (REscope) zooms into the failure regions and models the circuit performance distribution of likely-to-fail samples into a generalized pareto distribution (GPD), which is known as a good model of the tail of the PDF [16, 18]. It prunes the less useful process variation parameters in a high-dimensional problem by considering the contribution of each parameter to the performance metrics. Furthermore, we applied a nonlinear SVM classifier which is capable of identifying multiple disjoint failure regions. On a 108-dimension charge pump circuit in the PLL design, the proposed method outperforms the importance sampling approach and is 389x faster than the Monte Carlo approach. Moreover, it estimates the failure rate accurately, while importance sampling totally fails because the failure regions are not correctly captured.

The rest of this paper is organized as follows. In Section 2, the rare event modeling based on GPD are reviewed as the background of the proposed algorithm. In Section 3, we expatiate the proposed algorithm, including the algorithms performing parameter pruning, the nonlinear SVM classifier, and the algorithm that approximates the tail to GPD. Experiment results are presented in Section 4 to validate the accuracy and efficiency of proposed method. This paper is concluded in Section 5.

2. BACKGROUND
2.1 Rare Event Modeling

In statistical circuit simulation, a failure occurs event when a circuit performance metric does not meet the requirement. Mathematically, given a circuit with several process variation parameters $S = \{X_1, X_2, ..., X_N\}$, statistical circuit simulation analyzes the probability of circuit failure, i.e., a performance metric Y exceeds certain failure threshold y_f. The failure probability P_f can be represented as

$$P_f = P(Y > y_f) = 1 - F(y_f) \qquad (1)$$

where $F(y)$ is the cumulative distribution function (CDF) of performance metric Y.

A typical way to efficiently model Y is simulating a small size of Monte Carlo samples and applying moment matching to fit the simulation result into certain analytical form $F_{mm}(y)$ [6, 7, 8]. These approaches may correctly capture the overall shape of the distribution, it is, however, difficult to exactly fit the tail. The failure probability estimated by moment matching, $1 - F_{mm}(y_f)$, could be very inaccurate. Hence, we need to particularly model the tail of the distribution.

To simplify the discussion, let's assume that the perfor-
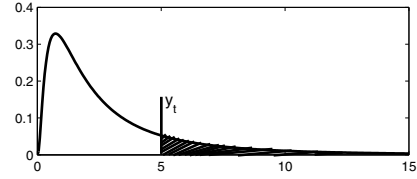
(a) Lognormal distribution
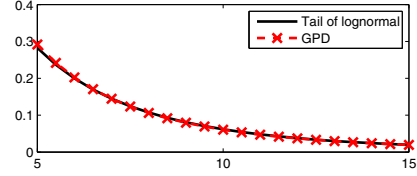
(b) Conditional PDF of the Lognormal tail

Figure 1: Model the tail of lognormal using GPD

mance metric, Y, belongs to a "lognormal" distribution, which is usually used to model circuit performance, i.e., memory read/write time. The PDF of a lognormal distribution is defined as

$$f_{\mu,\sigma}(y) = \frac{1}{y\sigma\sqrt{2\pi}} \exp(-\frac{(\ln y - \mu)^2}{2\sigma^2}) \qquad (2)$$

where μ and σ are the mean and standard deviation, respectively. A lognormal distribution with $\mu = \ln 2$ and $\sigma = 1$ is presented in Figure 1(a). Suppose y_t is a threshold that separates a tail from the body of the PDF function $f(y)$, the conditional CDF of the tail can be expressed as

$$F_t(y) = P(Y > y | Y > y_t) = \frac{F(y) - F(y_t)}{1 - F(y_t)} \qquad (3)$$

If $F(y_t)$ is known, the failure probability of the given threshold y_f can be calculated as:

$$P_f = (1 - F(y_t))(1 - F_t(y_f)) \qquad (4)$$

Fortunately, $F(y_t)$ can be accurately estimated by a few thousand samples because the event of $Y > y_t$ is not that rare. Therefore, the remaining problem is to correctly model the conditional CDF $F_t(y)$.

For several decades, the generalized pareto distribution (GPD) has been known as a good model for the distribution of the exceedence to a certain threshold in another distribution, i.e., the tail of $F(y)$ [18]. The CDF function of the GPD is defined as

$$F_{(\xi,\mu,\sigma)}(y) = \begin{cases} 1 - (1 - \frac{\xi(y-\mu)}{\sigma})^{\frac{1}{\xi}} & for\ \xi \neq 0 \\ 1 - \exp(-\frac{(y-\mu)}{\sigma}) & for\ \xi = 0 \end{cases} \qquad (5)$$

where ξ is the shape parameter, σ is the scale parameter, and μ is the starting point of the tail, i.e. y_t in this example. In particular, the tail of the lognormal random variable Y, can be accurately modeled by a GPD distribution with $\xi = 0.27$ and $\sigma = 3.5$, which is shown in Figure 1(b).

Given that the GPD can be used to model rare event, the remaining problems turn out to be 1) how to effectively draw samples in the tail to model the GPD in high dimensions, 2) how to deal with the problem of multiple failure regions, and 3) how to accurately fit the tail distribution into a GPD distribution.
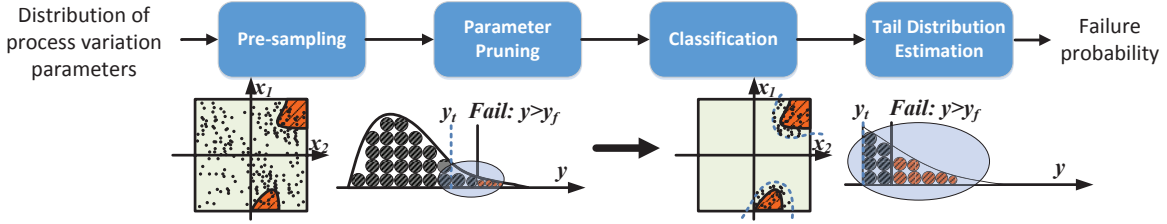
Figure 2: The REscope framework consists of four components: presampling, parameter pruning, classification, tail distribution estimation.

3. RARE EVENT MICROSCOPE
3.1 Algorithm overview

In this section, we present the proposed method, REscope, to identify multiple separate failure regions in high dimensional circuit simulation. REscope falls in the category of classification based methods. The REscope framework consists of four components, (1) presampling, (2) parameter pruning, (3)classification, and (4) tail distribution estimation, as shown in Fig. 2. REscope takes in the distribution of the process variation parameters, $S = \{X_1, X_2, ..., X_N\}$, of a test circuit, and outputs the estimated failure probability based on a given requirement on performance metric, i.e. $P_f = P(Y > y_f)$, where y_f is the threshold that determines circuit failure. In the remaining part of this section, we will elaborate on the design of each component in detail.

3.2 Presampling

The purpose of presampling is to approximately sketch the circuit behavior. Without loss of generality, we use M (typically a few thousand) Monte Carlo samples, $\mathcal{S} = \{s_1, s_2, ..., s_M\}$, subject to the distribution of S. Next, transistor level SPICE simulation is performed to evaluate the performance metric Y of the test circuit using these samples $\mathcal{S}$. A relaxed threshold y_t is chosen to determine the tail boundary from the main PDF, and probability that a sample falls in the tail $F(y_t) = P(Y > y_t)$ is calculated.

3.3 Parameter pruning

With up-scaling the design complexity and the advanced process technology, there are a sea of parameters in the circuit simulation. Parameter pruning, which maps the high-dimensional circuit description to a low-dimension space, can effectively improve the accuracy and efficiency of circuit simulation and analysis. Existing approaches, such as PCA, reduce the dimension by examining the correlation among input parameters and project them to a smaller, orthogonal base. However, PCA cannot help if each dimension of the process variation parameter, X_i and X_j, are modeled as mutually independent.

We leverage the ReliefF algorithm [19] to prune parameters in REscope. More specifically, each parameter is analyzed in terms of how sensitive it is to cause a circuit failure. The sensitivity is quantified as a weight parameter. In particular, for a data set $\mathcal{S} = \{s_1, s_2, ..., s_M\}$ with M samples, where each sample $s_i = \{x_1, x_2, ..., x_N\}$ consists of N variation parameters, ReliefF starts with a N-long weight vector ,W, of zeros, and iteratively updates W. In each iteration, it takes a random sample s_i, and finds the closest samples (in terms of Euclidean distance) in two decision regions respectively. The closest sample in the same region

is called "near-hit", and the other one is called "near-miss". The weight vector is then updated as

$$W_i = W_i + (x_i - nearMiss_i)^2 - (x_i - nearHit_i)^2 \quad (6)$$

The weight makes sense here because it increases if a feature differs from the nearby sample in the opposite region more than the sample in the same region, and decreases in the reverse case. It only requires $\mathcal{O}(NMlogM)$ time, and is noise-tolerant and robust to feature interactions.

Different from the general sensitivity analysis that only looks at the overall sensitively of the performance metric to a parameter, the ReliefF specifically looks at the sensitivity around the decision boundary of circuit pass and failure, which yields more important information than general sensitivity analysis.

3.4 Nonlinear SVM classifier

In the third step, a nonlinear classifier is adopted to identify whether a sample s_i falls in the unlikely-to-fail region or likely-to-fail region. Therefore, we can skip the unlikely-to-fail samples and focus on the samples in the tail.
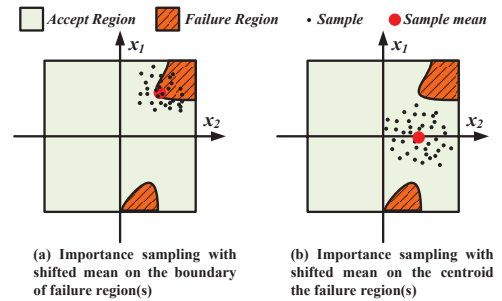

(a) Importance sampling with shifted mean on the boundary of failure region(s)

(b) Importance sampling with shifted mean on the centroid the failure region(s)

Figure 3: Importance sampling methods on a problem with two disconnected failure regions

A typical category of approaches for failure probability estimation is importance sampling. However, most of them are based on mean shifting and assume that only one failure region exists in the sample space. For example, [12] draws samples around the boundary of the failure region. While others, such as the HDIS [15], shifts the sample mean to the centroid of the failure regions. The "important samples" may easily cover all failure samples if there is only one failure region. In reality, there might be multiple disjoint failure regions. And the centroid of all failed samples might fall in somewhere outside the real failure regions, as shown in Figure 3.

A classification based approach considering the existence of multiple failure regions is proposed in [17]. The authors

assume that the samples in different failure regions yield different type of failures. Therefore, they applied the linear classifier multiple times to identify different failure types in a binary decision fashion. However, the assumption in [17] loses generality because it assumes that samples in different failure regions always yield different type of failures.

In REscope, we also consider a classification method to co-recognize the multiple failure regions. Different from [17], our method is not constrained to one failure type in one region, and is also applicable to the case with various failure types. Considering the intrinsic nonlinearity of circuit behavior, we employ a nonlinear classifier to tackle the multiple-region multiple-type failure sample classification challenges. More specifically, we use a Gaussian radial basis function kernel (GRBF) base support vector machine (SVM) to train and classify samples. The reason to choose GRBF kernel rather than linear or other polynomial kernel is that in high-dimensional circuits, the decision boundary between pass and failed samples is usually nonlinear. GRBF with radial arc boundary is more capable to adapt and discover the decision boundary.

3.5 Fitting the tail distribution to GPD

By performing classification, we can efficiently collect the likely-to-fail samples. Assuming $Y_p = \{y_{p1}, y_{p2}, ..., y_{pn}\}$ are the simulation outputs form the samples in the previous step that satisfy $y_{pi} > y_t$, this step approximates the distribution of Y_p to a GPD. As given in equation (5), there are only three parameters, ξ, μ, and σ, to determine the CDF of GPD. While in this example, only ξ and σ need to be approximate since the parameter μ is known as the start point of the tail, y_t. There are three approaches to approximate ξ and σ in the CDF, moment matching [18], probability-weighted moment (PWM) matching [20], and maximum likelihood estimation (MLE) [21].

The moment matching and PWM matching only use the first two order of moments to estimate these two parameters, which may lead to a mismatch in high order statistics. On the other hand, the MLE iteratively approaches the $\hat{\xi}$ and $\hat{\sigma}$ using Newton's method towards a maximum log likelihood function [21]:

$$\log L(Y_p; \xi, \sigma) = -n \log(\sigma) - (1 - \xi) \sum_{i=1}^{n} z_i \qquad (7)$$

where $z_i = -\xi^{-1} \log(1 - \xi y_{pi}/\sigma)$.

The drawback of MLE is that it may take many iterations before the results finally converge to $\hat{\xi}$ and $\hat{\sigma}$, which maximize the log likelihood function [21]. In REscope, we use PWM matching results, ξ_0 and σ_0, as the initial solution of the Newton's method. Next, MLE is applied to iteratively approach the $\hat{\xi}$ and $\hat{\sigma}$. The number of iterations is reduced because of the accurate and non-arbitrary starting point.

4. EXPERIMENT RESULTS
4.1 Charge pump circuit and experiment setting

The performance of REscope is evaluated using a charge pump (CP) circuit, which is a critical sub-circuit of the phase-locked loop (PLL). The block diagram of a PLL is presented in Figure 4.

As a sub-circuit of the PLL, CP adjusts the frequency of the output clock signal, CLK_{out}, via a charge/discharge
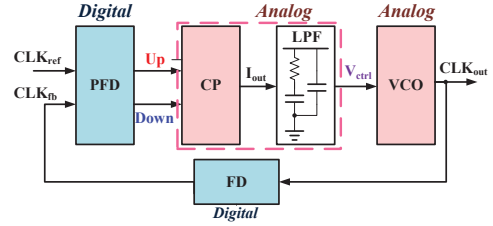

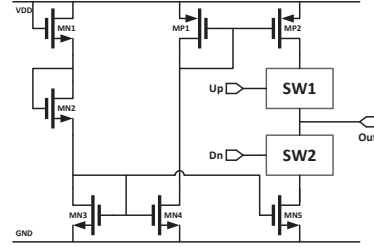**Figure 4: A block diagram of PLL**

Figure 5: Simplified schematic of the charge pump circuit

capacitor and voltage controlled oscillator (VCO). A simplified schematic of the charge pump consisting of two switched current sources is illustrated in Figure 5. If the output clock CLK_{out} lags behind the input reference clock, CLK_{ref}, the up signal will be high and down signal will be low. The up signal turns on the upper switch and charges the output node. On the other hand, when CLK_{out} leads CLK_{ref}, the up/down signal controls the switches to generate a discharge current at the output node. Finally, when CLK_{ref} and CLK_{out} are synchronized in frequency and phase, both up signal and down signal are set as low, leading to a zero net current I_{out} and a constant V_{ctrl}.

The CP circuit acts as a critical module in the entire control loop. However, when there is a mismatch between transistor MP2 and MN5 in Figure 5, the net current at the output node is not zero. It could cause large fluctuation at the control voltage, also known as "jitter", which may critically effect the system stability. In this work, we define a failure as the mismatch of charge current and discharge current, mathematically, $\max(\frac{I_{Charge}}{I_{Discharge}}, \frac{I_{Charge}}{I_{Discharge}}) > \gamma$, where γ is a threshold of this performance matric.

In the experiment, we designed the CP circuit using TSMC 45nm technology and simulated it with HSPICE with BSIM4 transistor model. In each transistor, we consider 4 parameters, channel-length offset (δL), channel-width offset (δW), gate oxide thickness (t_{ox}), and threshold voltage (V_{th}), as the source of process variations as suggested by the foundry.

REscope is used to evaluate the mismatch current of the CP circuit in Figure 5. In addition, Monte Carlo (MC), statistical blockade (SB) [1] have been implemented, and the source code of HDIS [15] was obtained from its authors for accuracy and efficiency comparison. We evaluate the efficiency by counting the total number of simulations required to yield a stable failure rate. In REscope, we generate a large number of MC samples and filter them by the classifier to make sure we can get enough samples in the tail. In our implementation, REscope stops when there are 1000 samples fall on the tail. MC converges when the relative

standard deviation of the failure probability, $\sigma_r = \frac{std(p_f)}{p_f}$, is smaller than 0.1.

4.2 Handling multiple separate failure regions

The CP is a typical circuit with multiple failure regions. To illustrate the capability of REscope in handling multiple failure regions, we use a simplified process variation model, which only considers the threshold voltage (V_{th}) of MP2 and MN5 in Figure 5 as the source of process variations. When the V_{th} of MN5 is lower than the nominal value and V_{th} of MP2 is higher than the nominal, there will be a mismatch as $I_{Discharge}$ can be larger than I_{Charge}, and vice verse.

In this experiment, the threshold y_t is configured to ensure a 5% failure rate. Under these configurations, the failure regions can be clearly visualized on a 2-D space, as shown in Figure 6(a).
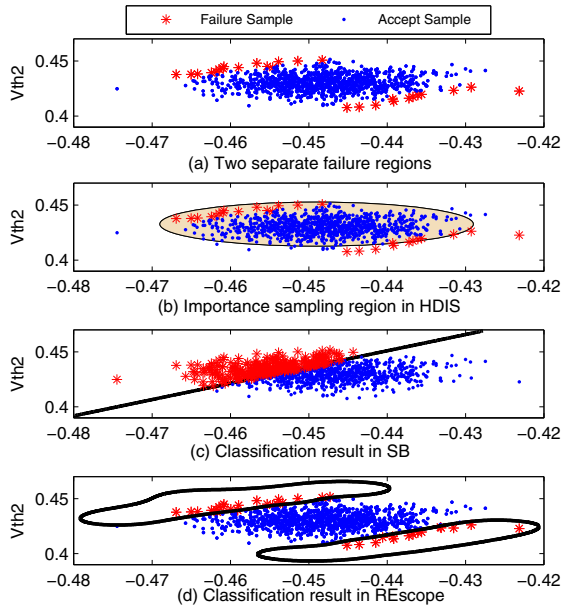
Figure 6: How multiple failure regions are handled in HDIS [15], SB [16], and REscope

The importance sampling region of HDIS, along with the classification results of SB and REscope are illustrated in Figure 6(b), (c), and (d), respectively.

It is easy to notice that HDIS fails to effectively capture the "important" samples, because it attempts to draw samples around the centroid of the failure region. When there are 2 failure regions, as illustrated in Figure 6(b), however, the centroid falls almost in the center of a *success* region, which make it difficult to cover truly "important" samples.

SB adopts a linear classifier, which essentially finds a linear hyperplane separating the successful region and fail region. However, in this example, it is impossible to separate all failed samples from the successful ones using just a linear hyperplane. In Figure 6(c), SB draws a boundary cutting through the successful region, which introduces a lot of over-classified samples on the top-left sample space. Moreover, it only covers the failed samples on the top-left corner and mis-classifies all the failed samples on the bottom-right corner of the sample space.

By taking advantage of the nonlinear classifier, REscope successfully classifies all failed samples in the sample space, which is illustrated in Figure 6(d).

4.3 Parameter pruning

In the following discussion, we model the δL, δW, t_{ox}, and V_{th} in all 27 transistors of the charge pump circuit as process variation source, and evaluate the current mismatch.
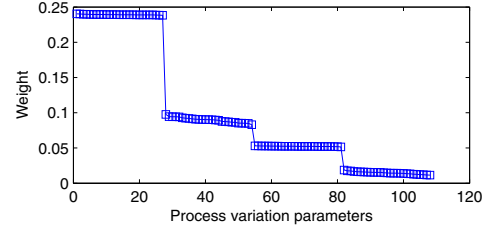
Figure 7: Weight of all 108 process variations in charge pump circuit

On this 108-dim problem, ReliefF is performed to reduce the dimension before constructing the classifier. For each process variation parameter, the weight is evaluated and their weights are ranked and illustrated in Figure 7. It is easy to notice that the maximal weight can be more than 10x greater than the minimal one.

A high number of input variables will require correspondingly larger number of samples to train the classifier, which is not efficient in practice. On the other hand, a high input variable dimension with small number of samples may fool the classifier. In practice, we normalize the weights and setup a threshold to prune the parameters with smaller weights than the threshold. In this example, we only kept the first 27 parameters and used them to build the classifier.

4.4 Accuracy and Efficiency

On this 108-dim problem, REscope is compared with MC and HDIS on efficiency and accuracy. The results are presented in Table 1. SB is excluded from the comparison because the linear classifier generated by SB accepts all the MC samples, which makes no difference between it and MC simulation. HDIS outputs a nearly random failure probability with 20 thousand simulations since it fails to shift the mean to a desired place. REscope accurately calculates the failure probability as 2.256e-5, with only 1.05% relative error compared with MC.

On the efficiency side, MC needs 1.4 million samples to reach a confident estimation of the failure probability 2.279e-5, which is around 4.07 sigma. Beyond, 4.07 sigma, the MC result may be unreliable. On the other hand, REscope only requires 2000 samples to construct the nonlinear classifier. Next, 100,000 MC samples are generated for evaluation, but only 1621 are actually simulated, including 630 over-classified samples to avoid misclassification. Therefore, REscope achieves 389x speedup compared to MC almost without sacrificing accuracy.

To examine how the approximation accuracy scales when the failure probability becomes more rare, we plot the CDF tail of mismatch current estimated by REscope in Figure 8(a), which perfectly matches to the MC result. In Figure 8(b), the fitting results are illustrated more clearly after we represent the CDF in terms of sigma (in log scale). The

Table 1: Comparison of the accuracy and efficiency on charge pump circuit

	Monte Carlo (MC)	Importance sampling (HDIS)[15]	Proposed approach (REscope)
failure probability	2.279e-5 (0%)	1.136e-3	2.256e-5 (+1.05%)
#sim. runs	1.4e+6 (389x)	2e+4 (5.6x)	3.6e+3 (1x)
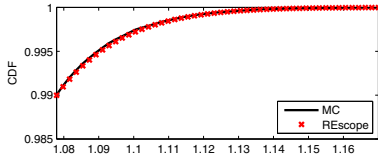
(a) Tail of CDF in linear scale
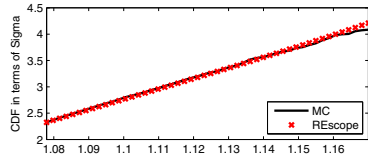
(b) Tail of CDF in log scale

Figure 8: Modeling the tail of the mismatch current distribution

REscope estimates the probability of rare event accurately for upto 4.2 sigma, which is about 1.22e-5 in terms of probability. Beyond 4.2 sigma, MC cannot guarantee the accuracy as only 1.4 million MC samples are available.

5. CONCLUSION

In this paper, REscope is proposed for statistical circuit simulation with rare failure event. Given a circuit with a large number of process variation parameters, REscope first leverages the ReliefF algorithm to evaluate each parameter, and prune those that have little contribution to the circuit failure. Furthermore, we applied a nonlinear classifier which is capable of identifying disjoint multiple failure regions. Because of the classification, the computation complexity is reduced by only simulating samples that are classified as likely-to-fail samples. When sufficient samples are simulated, the simulation results are approximated to a GPD, which is usually used model the rare event. On a 108-dimension charge pump circuit, the proposed method outperforms the importance sampling approach and is more than 2 orders faster than the Monte Carlo approach. Moreover, it estimates the failure rate accurately, while importance sampling totally fails because the failure regions are not correctly captured.

6. REFERENCES

[1] A. Singhee and R. A. Rutenbar, "Statistical blockade: very fast statistical simulation and modeling of rare circuit events and its application to memory design," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 28, no. 8, pp. 1176–1189, 2009.

[2] C. Jacoboni and P. Lugli, *The Monte Carlo method for semiconductor device simulation*. Springer, 1989, vol. 3.

[3] W. Wu, Y. Shan, X. Chen, Y. Wang, and H. Yang, "Fpga accelerated parallel sparse matrix factorization for circuit simulations," in *Reconfigurable Computing: Architectures, Tools and Applications*. Springer, 2011, pp. 302–315.

[4] W. Wu, F. Gong, R. Krishnan, L. He, and H. Yu, "Exploiting parallelism by data dependency elimination: A case study of circuit simulation algorithms," *Design Test, IEEE*, vol. 30, no. 1, pp. 26–35, Feb 2013.

[5] X. Chen, W. Wu, Y. Wang, H. Yu, and H. Yang, "An escheduler-based data dependence analysis and task scheduling for parallel circuit simulation," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 58, no. 10, pp. 702 –706, oct. 2011.

[6] X. Li, J. Le, P. Gopalakrishnan, and L. T. Pileggi, "Asymptotic probability extraction for nonnormal performance distributions," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 26, no. 1, pp. 16–37, 2007.

[7] F. Gong, H. Yu, and L. He, "Stochastic analog circuit behavior modeling by point estimation method," in *Proceedings of the 2011 international symposium on Physical design*. ACM, 2011, pp. 175–182.

[8] R. Krishnan, W. Wu, F. Gong, and L. He, "Stochastic behavioral modeling of analog/mixed-signal circuits by maximizing entropy." in *ISQED*, 2013, pp. 572–579.

[9] P. Feldmann and R. W. Freund, "Efficient linear circuit analysis by padé approximation via the lanczos process," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 14, no. 5, pp. 639–649, 1995.

[10] E. Chiprout and M. Nakhla, *Asymptotic waveform evaluation and moment matching for interconnect analysis*. Kluwer Academic Publishers, 1994.

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

[12] L. Dolecek, M. Qazi, D. Shah, and A. Chandrakasan, "Breaking the simulation barrier: SRAM evaluation through norm minimization," in *Proceedings of the 2008 IEEE/ACM International Conference on Computer-Aided Design*, ser. ICCAD '08, 2008, pp. 322–329.

[13] M. Qazi, M. Tikekar, L. Dolecek, D. Shah, and A. Chandrakasan, "Loop flattening and spherical sampling: Highly efficient model reduction techniques for SRAM yield analysis," in *Design, Automation Test in Europe Conference Exhibition (DATE), 2010*, 2010, pp. 801 –806.

[14] K. Katayama, S. Hagiwara, H. Tsutsui, H. Ochi, and T. Sato, "Sequential importance sampling for low-probability and high-dimensional SRAM yield analysis," in *IEEE/ACM International Conference on Computer-Aided Design*, 2010.

[15] W. Wu, F. Gong, G. Chen, and L. He, "A fast and provably bounded failure analysis of memory circuits in high dimensions," in *19th Asia and South Pacific Design Automation Conference (ASP-DAC)*, 2014, pp. 424–429.

[16] A. Singhee and R. A. Rutenbar, "Statistical blockade: a novel method for very fast monte carlo simulation of rare circuit events, and its application," in *Design, Automation, and Test in Europe*, 2008, pp. 235–251.

[17] A. Singhee, J. Wang, B. H. Calhoun, and R. A. Rutenbar, "Recursive statistical blockade: an enhanced technique for rare event simulation with application to sram circuit design," in *21st International Conference on VLSI Design*. IEEE, 2008, pp. 131–136.

[18] J. R. Hosking and J. R. Wallis, "Parameter and quantile estimation for the generalized pareto distribution," *Technometrics*, vol. 29, no. 3, pp. 339–349, 1987.

[19] I. Kononenko, E. Šimec, and M. Robnik-Šikonja, "Overcoming the myopia of inductive learning algorithms with RELIEFF," *Applied Intelligence*, vol. 7, no. 1, pp. 39–55, 1997.

[20] J. Hosking, J. R. Wallis, and E. F. Wood, "Estimation of the generalized extreme-value distribution by the method of probability-weighted moments," *Technometrics*, vol. 27, no. 3, pp. 251–261, 1985.

[21] J. Hosking, "Algorithm as 215: Maximum-likelihood estimation of the parameters of the generalized extreme-value distribution," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 34, no. 3, pp. 301–310, 1985.