Cross Entropy Minimization for Efficient Estimation of SRAM Failure Rate

Mohammed Abdul Shahid
Electrical Engineering Department,
University of California, Los Angeles, CA 90095, USA
Email: amohammed@ucla.edu

***Abstract*—As the semiconductor technology scales down to 45nm and below, process variations have a profound effect on SRAM cells and an urgent need is to develop fast statistical tools which can accurately estimate the extremely small failure probability of SRAM cells. In this paper, we adopt the Importance Sampling (IS) based information theory inspired *Minimum Cross Entropy* method, to propose a general technique to quickly evaluate the failure probability of SRAM cells. In particular, we first mathematically formulate the failure of SRAM cells such that the concept of 'Cross Entropy Distance' can be leveraged, and the distance between the *ideal distribution* for IS and the *practical distribution* for IS (which is used for generating samples), is well-defined. This cross entropy distance is now minimized resulting in a simple analytical solution to obtain the *optimal practical distribution* for IS, thereby expediting the convergence of estimation. The experimental results of a commercial 45nm SRAM cell demonstrate that for the same accuracy, the proposed method yields computational savings on the order of 17$\sim$50X over the existing state-of-the-art techniques.**

I. Introduction

The current fabrication technology in the nanometer regime, poses a number of new challenges due to the variations and inaccuracies in the manufacturing processes. These process variations (e.g., random doping fluctuations) have a substantial impact on bitcells in memory designs, as these components are replicated in large arrays, and to produce even a single working product requires that a large number of these replicated components all work correctly. This translates to a tolerable *rare* failure of one or two defects in hundreds of millions or billions of instances of these components, thus making it inherently difficult to not only design these components [1] but also verify them. Henceforth, the effect of process variations on contending design specifications of SRAM bitcells [2], has created the immediate need for fast and extremely accurate estimations of their failure probability in order to achieve comprehensive verifications and guarantee robust designs.

In the recent past, a number of approaches based on statistical analysis have been proposed and investigated for verification of SRAM circuits and their associated *rare* failure events [3]–[10]. The most common and traditional approaches have been based on Monte Carlo (MC) ([3]-[5], [7] and [9]) or designing analytical performance measurement models to predict the SRAM parametric yield ([1] and [8]) or a combination of both the approaches [10]. Inspite of attributes like quantifiable accuracy and scalability that is independent of dimensionality, a forthright and unsophisticated implementation of MC requires a prohibitive amount of time (100's of millions, or billions of samples in order to produce a handful of failures) to obtain accurate information. The accuracy of the analytical models and their ability to exactly capture the circuit behaviour is also a matter of concern. Thus, most of the traditional approaches fail to quickly estimate the extreme statistics of rare events of SRAM circuits.

The more recent approaches to improve the sampling efficiency of MC have been based on Importance Sampling (IS) [3]-[5], [7]-[9]. In IS, the original distribution (PDF) is *shifted* towards the rare infeasible failure region, and using this new shifted distribution called *practical distribution* for IS, the failure region is now directly sampled. The non-triviality lies in determining the *optimal shift*, which results in the *optimal practical distribution* for IS, in order to predict quickly and accurately. In [3]-[5], the *optimal practical distribution* for IS, is found using a numerical optimization method called *norm minimization* and for high dimensional problems a variant of it called *spherical sampling* was proposed. However, more recently, in [7] and [9], these methods have been shown to have sub-optimal performance mainly due to the use of sub-optimal shift and their performance degradation for high dimensional problems.

In [7], the optimal distribution is found using Gibbs Sampling which although provides improved results when compared to [3]-[5], suffers from issues like performing complex and costly binary search to generate even a single Gibbs sample, which only gets difficult for high dimensional problems. In [9], particle filtering is used which again involves complex procedures with many approximations whose best results are impacted by the choice of parameters. In [6], MC samples are drawn, and a classifier is used to "block" those MC samples which are not in the 97th percentile tails and the remaining samples are simulated. The ones beyond the 99th percentile are used to construct the tail distribution. However, apart from the significant challenges in training the classifier there is also no way to guarantee the efficiency of the classifier and the possible errors which might occur due to its inefficiency. Furthermore, the efficiency of this method for high dimensional problems is yet to to be demonstrated.

Our Contributions: In this work, we adopt the Minimum Cross Entropy (MCE) method from the statistics community to develop a *complete* and *general* statistical methodology called ***Cross Entropy Minimization for SRAM***, to give quick and

978-3-9810801-8-6/DATE12/©2012 EDAA

accurate evaluation of SRAM designs. In particular, we first mathematically formulate the failure of SRAM cells in a way such that IS can be applied to it, and the MCE method is then applied to determine the *optimal practical distribution* for IS. A number of implementation issues like the initial vector selection and the numerical optimization problem to determine the *optimal shift* are well taken care of, and in fact a simple analytical solution is derived for the same. The proposed method is illustrated using the example of Data Retention Voltage (DRV) and the computational speed up and increased accuracy of it are compared to the existing state-of-the-art techniques like Spherical Sampling (SS) and Mixture Importance Sampling (MixIS). Inspite of being an IS-based approach, the proposed method can be extended to handle high dimensional problems but due to the page limit of the paper, the application has not been discussed and has been reserved for a follow-up paper. To the best of our knowledge, this is the first time such a methodology has been applied for SRAM failure analysis.

Organization: The remainder of this paper is organized as follows: In Section II, the mathematical formulation of SRAM failure and the necessary background on MCE is provided. Section III contains the detailed description of the proposed method to evaluate the SRAM failure. The applicability, accuracy and significant computational speed-up of the proposed method are demonstrated through an example provided in Section IV. Section V summarizes the main contributions of this paper and proposes future extensions.

II. Background on SRAM Failure and Minimum Cross Entropy Method

In this section, we first present the mathematical formulation for SRAM failure and then briefly recapitulate the Minimum Cross Entropy (MCE) Method [11], and formulate it conceptually and mathematically in a way that is suitable and easy to be applied to estimate the probability of failure of SRAM circuits.

A. Mathematical Formulation of SRAM failure

Here we build on the mathematical formulation of SRAM failure presented in [4] and make it suitable for the application of MCE. Let A be the rare event of failure of an SRAM cell which is characterized by a random variable X. This rare event could be the Static Noise Margin (SNM) being less than zero or the read/write failure. Usually, this random variable, X, is a nonlinear and complicated function, F, of circuit variables (threshold voltages or channel length of transistors) which under process variations, deviate from their nominal values by a significant amount. Without loss of generality, we can define $X = F(Y_1, ..., Y_M)$ where Y_i represents the circuit variables. The deviations in the circuit variables are difficult to estimate which make it increasingly difficult to determine the distribution of the random variable X, and it usually has a long tail distribution. In this paper, the deviations in the circuit variables are modelled as independent Gaussian random variables with mean u_i and variance σ_i^2 and their distribution can be represented as $f(y_i, u_i)$ i.e., $Y_i \sim N(u_i, \sigma_i^2)$. The modelling is justified as the process variations are random and their effect on the circuit variables is largely independent [4]. These distributions are used to generate samples and based on the fraction of samples which result in occurance of event A, the probability of failure, p_{fail}, is estimated.

In Importance Sampling, the original distribution is shifted to the rare infeasible failure region and specifically for Gaussian distributions, the mean u_i is *shifted* to a new mean v_i with variance remaining the same. Based on this shift, the new Gaussian distribution $\hat{f}(y_i, v_i)$ i.e., $\hat{Y}_i \sim N(v_i, \sigma_i^2)$ is determined which is now used to generate samples and make an estimation of p_{fail}. The shifted distribution which is used to generate samples is called the *practical distribution* for IS and to achieve maximum speed up the *optimal practical distribution* for IS is to be determined. In order to make an estimation of p_{fail} using IS, we make use of the weight function $w(\hat{y}, u, v)$, where

$$w(\hat{y}, u, v) = \frac{\prod_{i=1}^{M} f(\hat{y}_i, u_i)}{\prod_{i=1}^{M} \hat{f}(\hat{y}_i, v_i)}, \text{for all } \hat{y}, \tag{1}$$

and for N independent samples derived from $\hat{f}(\hat{y}, v)$, the estimation of p_{fail} is

$$p_{fail} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(\hat{y}^{(i)} \in A) w(\hat{y}^{(i)}, u, v), \tag{2}$$

where the truth-indicator $1(ST)$ takes value 1 if ST is true, and 0 if ST is false. For the specific case of Gaussian distribution as shown in [4], the weight function, $w(.)$, can be calculated as follows,

$$w(\hat{y}, u, v) = \frac{exp\left(-\sum_{i=1}^{M} \frac{(\hat{y}_i - u_i)^2}{2\sigma_i^2}\right)}{exp\left(-\sum_{i=1}^{M} \frac{(\hat{y}_i - v_i)^2}{2\sigma_i^2}\right)}, \tag{3}$$

$$= exp\left(-\sum_{i=1}^{M} \frac{2\hat{y}_i(v_i - u_i) - (v_i^2 - u_i^2)}{2\sigma_i^2}\right) \tag{4}$$

The above expression can now be used in equation (2) to get an estimate of p_{fail}. The *optimum* choice of the new shifted mean v, say v^* to determine the *optimal practical distribution* for IS plays a major role in quick convergence of the estimations. Further details on determining this is provided in the next section.

B. Minimum Cross Entropy-General Methodology

In this section, we introduce the MCE method [11] which has been widely used in various statistical inferences and machine learning problems. The MCE Method basically finds the best possible change of measure, *optimal practical distribution* for IS, $\hat{f}(y, v^*)$, which is *closest in distance* to *ideal distribution* for IS, $f^{ideal}(y)$.

It is well-known and has also been pointed out in [7] that if $f(y, u)$ is the original distribution of Y then the *ideal*

distribution for IS which has the *ideal* shifted mean, is given by

$$f^{ideal}(y) = \frac{\mathbf{1}(y \in A) f(y,u)}{p_{fail}}. \quad (5)$$

However, we cannot make use of the above equation to get the optimum shift, as the indicator function is not known in advance and more importantly we have an unknown p_{fail}. So, the idea is to find some distribution, *optimal practical distribution* for IS, $\hat{f}(y, v^*)$, which in some sense is closest to the *ideal* distribution of equation (5) , $f^{ideal}(y)$. This *sense* of finding a distribution which is ***closest in distance*** to $f^{ideal}(y)$ can be borrowed from *information theory*.

The MCE method minimizes the **Kullback-Leibler (KL-Divergence)** or **Cross Entropy (CE)** distance between $f^{ideal}(y)$ and $\hat{f}(y, v^*)$. The cross entropy distance between any two distributions is defined in [12] as

$$D = E_{f^{ideal}}[log(\frac{f^{ideal}(y)}{\hat{f}(y, v^*)})], \quad (6)$$

where $E[.]$ is the expectation operator. We are interested in minimizing the above defined cross entropy distance in order to get the distribution which is closest in distance to the *ideal* distribution i.e.,

$$\textit{Minimize } D = E_{f^{ideal}}[log(\frac{f^{ideal}(y)}{\hat{f}(y, v^*)})]. \quad (7)$$

As shown in [11], it is not difficult to see that the above equation reduces to the following numerical optimization expression

$$v^* = \arg\max_{v} E_v[\mathbf{1}(\hat{y} \in A) log(\hat{f}(\hat{y}, v))]. \quad (8)$$

The above expression gives the optimum shifted mean vector, v^* of the distribution $\hat{f}(y, v^*)$ (*optimal practical distribution* for IS) which is closest in distance to the *ideal* distribution for IS, $f^{ideal}(y)$.

However, there are two issues with equation (8) and they are as follows

- The indicator function $\mathbf{1}(\hat{y} \in A)$ is an indicator of a rare event and we may not get enough samples to derive the optimum shifted mean vector, v^*.
- It is a numerical optimization problem and though conceptually it is closest to the ideal distribution, it is computationally expensive to calculate it.

The two issues are taken care of in the next section.

III. Cross Entropy Minimization for SRAM

So far, it has been established that to achieve best performance using IS, the optimal mean shift vector is to be determined and it is shown in the previous section it can be found using MCE (equation (8)). In this section, we take care of all the issues pointed out in the previous section and establish the proposed approach, ***Cross Entropy Minimization for SRAM***.

A. Multi Level Minimum Cross Entropy

It can be observed from equation (8) that to derive the optimum shifted mean vector we need to make use of the indicator function $\mathbf{1}(\hat{y} \in A)$. However, the indicator function is an indicator of a rare event and we may not get enough samples to derive the optimum shifted mean vector. An extension to the MCE method to address this issue is to make use of *multiple levels* of cross entropy method as in [11], wherein the original distribution $f(y, u)$ is initially shifted to some approximate distribution $\hat{f}(y, l)$ and this shifting would result in the weight function $w(\hat{y}, u, l)$. The initial shifted mean vector, l , should be chosen so that the event becomes more likely under this new shifted distribution i.e., it is a *less* rare event. Making use of *multiple levels* of cross entropy method, equation (8) can now be modified as

$$v^* = \arg\max_{v} E_v[\mathbf{1}(\hat{y} \in A) w(\hat{y}, u, l) log(\hat{f}(\hat{y}, v))]. \quad (9)$$

This still leaves us with the important decision of the choice of this vector l. We propose to find this vector using the state-of-the-art *norm-minimization* approach.

B. Initial Vector Selection

The following step-by-step procedure can be followed to calculate the vector l.

1) Generate a few uniform random shifts (*say* 100) of the original means u_i to get the new shifted means l_i, for each of the circuit variables from their respective variation spaces.
2) Use these shifted means to obtain the appropriate shifted distributions, $\hat{f}(y, l_i)$, generate the samples using them, run the simulations and screen out those shifted means which resulted in a failure event.
3) These screened out shifted means are now used and their L2-norm is calculated using the following equation

$$\textit{L2-Norm of } l = \sum_{i=1}^{M} \frac{(l_i - u_i)^2}{2\sigma_i^2}. \quad (10)$$

4) A good choice for the inital vector would now be the shifted mean vector which resulted in the minimum value of L2-Norm i.e.,

$$\textit{Initial vector } l = \arg\min_{l} \sum_{i=1}^{M} \frac{(l_i - u_i)^2}{2\sigma_i^2}. \quad (11)$$

It should be noted that the idea is to just get an approximate initial direction to make the event less rare and not solve it completely and accurately. This still leaves us with the second and more important issue of solving the equation (9), a numerical optimization problem, to determine the optimum shifted mean vector. It however turns out that for certain distributions especially the exponential families (which is of great interest for most of the SRAM analysis problems), the expression breaks down to a simple analytical expression as will be shown next.

C. Minimum Cross Entropy and the Natural Exponential Families

If we consider a univariate family of distributions with densities (pdf's) $f_\theta, \theta \in \Theta$. The family is said to be a Natural Exponential Family (NEF) as defined in [13]

$$f_\theta(y) = e^{(y\theta - \kappa(\theta))} h(y). \tag{12}$$

Now, every NEF distribution with pdf f_0 for which the moment generating function exists in a neighbourhood of **0** generates its own NEF by letting κ be the cumulant function

$$\kappa(\theta) = log \int e^{y\theta} f_0(y) dy, \tag{13}$$

and by defining

$$f_\theta(y) = e^{(y\theta - \kappa(\theta))} f_0(y). \tag{14}$$

Based on the above equation we can say that f_θ is obtained from f_0 by an exponential twist/tilt with twisting/tilting parameter θ.

Now, the circuit variables, Y, are modeled as Gaussian Variables (which belongs to the exponential family) and thus Y will have a distribution in the NEF f_θ. Then as observed in [11], it is not difficult to see that

$$u = mean = E_\theta[Y] = \kappa^{'}(\theta), \tag{15}$$

and

$$\sigma^2 = Variance = \mathrm{Var}_\theta(Y) = \kappa^{''}(\theta). \tag{16}$$

Since $\kappa^{'}$ is increasing we may reparametrize the family using the mean u. In particular, to the NEF above corresponds a family g_u such that for each pair (θ, u) satisfying $\kappa^{'}(\theta) = u$ we have $g_u = f_\theta$. Now, the circuit variable Y is a random variable from NEF and it can be reparametrized by the mean u as follows

$$f(y, u) = f_{\theta(u)}(y) = g_u(y) = exp(\theta(u)y - \kappa(\theta(u)))h(y). \tag{17}$$

In equation (9), we wish to maximize with respect to v^* the function D defined as

$$D(v^*) = E_v[\mathbf{1}(\hat{y} \in A) w(\hat{y}, u, l) log(\hat{f}(\hat{y}, v))], \tag{18}$$

where $\hat{f}(\hat{y}, v)$ is now of the form as in equation (17). Solving for $D^{'}(v^*) = 0$, and converting it in its sampled form, we get

$$v^* = \frac{\sum_{i=1}^{N} \mathbf{1}(\hat{y}^{(i)} \in A) w(\hat{y}^{(i)}, u, l) \hat{y}^{(i)}}{\sum_{i=1}^{N} \mathbf{1}(\hat{y}^{(i)} \in A) w(\hat{y}^{(i)}, u, l)}. \tag{19}$$

The above is a simple analytical expression to calculate the optimum mean shifted vector for the distribution $\hat{f}(y, v^*)$. We can now present the algorithm for the proposed method.

D. Algorithm: Cross Entropy Minimization for SRAM

1) If the original distribution is $f(y, u)$, choose a particular l , as demonstrated in the Section III.B, so that the event becomes *'less rare'*.
2) Generate N_1 samples from the shifted distribution $\hat{f}(y, l)$.
3) Calculate the optimum shifted mean vector, v^*, using equation (19).
4) Use this optimum shifted mean vector , v^*, in the distribution $\hat{f}(y, v^*)$. This is the *optimum practical distribution* for IS based on ***Cross Entropy Minimization for SRAM***.
5) Generate N samples from the *optimum practical distribution*, $\hat{f}(y, v^*)$.
6) The estimate of p_{fail}, using importance sampling is now given by

$$\hat{p}_{fail} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(\hat{y}^{(i)} \in A) w(\hat{y}^{(i)}, u, v^*), \tag{20}$$

where $w(\hat{y}^{(i)}, u, v^*)$ is given using equation (4). Thus, the proposed method is general, complete and simple with an analytical solution. The experimental results and superior performance of the proposed method when compared to MC and other techniques is demonstrated in the next section.

IV. Experimental Results

In this section, the proposed, ***Cross Entropy Minimization for SRAM***, is illustrated by means of an example. As such the proposed method is general and can be used for any of the statistical circuit analysis problems but we focus on the prediction of failure rate of SRAMs, as the maximum impact of process variations is seen in them. The results demonstrate the accuracy, significant computational gains and efficiency of the proposed approach when compared to other state-of-the-art techniques.

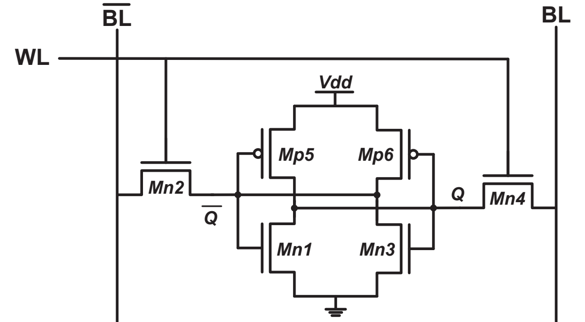


Fig. 1: The schematic of the $6T$ SRAM cell

A. Experimental Set-up

The most effective approach for leakage power savings during SRAM standby mode is the scaling of Supply voltage (V_{DD}). However, lowering V_{DD} as far as possible although maximizes the reduction in leakage power but also leads to

TABLE I: Figure of Merit Comparison for $1E+4$ Samples

	MC	MixIS	SS	Proposed
p_{fail}	5.455E-4	3.681E-4	4.342E-4	4.327E-4
ρ	0.8129	0.1111	0.9831	0.022
Accuracy	18.71%	88.53%	90.42%	97.8%
# of runs	1.0E+4	1.0E+4	1.0E+4	1.0E+4

data loss. The data retention voltage (DRV) is the *lower bound* of the standby supply voltage that still preserves the data in the SRAM cells [4]. The example presented here considers the data retention voltage (DRV) for a $6T$ SRAM (shown in Figure 1) designed in a commercial 45nm process. Specifically, the issue is how low can the supply voltage, V_{DD}, be set such that the original data is preserved in the SRAM cell [4] and to identify it we ran the simulation for three different V_{DD} levels.

For this problem, there are $M = 6$ circuit variables i.e., threshold voltages of the transistors and they are modelled as independent Gaussian random variables, $f(y_i, u_i)$, with mean as the nominal threshold voltage ($0.466V$ for PMOS and $-0.4118V$ for NMOS) and variances as 10% of the nominal threshold voltages. The rare event of interest is the failure, Static Noise Margin (SNM) ≤ 0. For evaluation and comparison purposes, the following approaches are implemented: 1) Standard Monte Carlo (MC), 2) Mixture Importance Sampling (MixIS) [3], 3) Spherical Sampling (SS) [5] and 4) the proposed, ***Cross Entropy Minimization for SRAM***.

B. Efficiency of Cross Entropy Minimization for SRAM

To compare the computational savings of all the approaches, the *figure of merit*, ρ, as defined in [4] is calculated. For all the approaches, we ran simulations until we could atleast achieve a figure of merit of $\rho = 0.1$ i.e., 90% accuracy with 90% confidence. The experimental results are presented next and based on a closer analysis of them several important observations are made.

1) Accuracy: To validate the accuracy of proposed method and other state-of-the-art existing techniques, we compare their failure probability estimates with the standard Monte Carlo for different V_{DD} levels. It can be observed from Table-II, that the value of p_{fail} for all the techniques, for different V_{DD} levels, are in close agreement with each other. Specifically, we can observe from Fig.(2(a)) plotted for $V_{DD} = 300mV$ that, the final predictions of p_{fail} for all the techniques are extremely close to each other thereby confirming the accuracy of the estimates of all the previous techniques and our proposed approach. In addition, the proposed method starts with an estimate of p_{fail} which is very close to the final estimation of p_{fail}, which points out to the optimality of the mean shifted vector for IS derived by the proposed approach. This explains the quick convergence achieved by the proposed approach.

2) Computational Savings: The proposed approach offers significant computational savings when compared to the other techniques and based on the experimental results presented in Table-II and shown in the plot of the evolution of the figure of merit, ρ, (see Fig.2(b)), it can be observed that

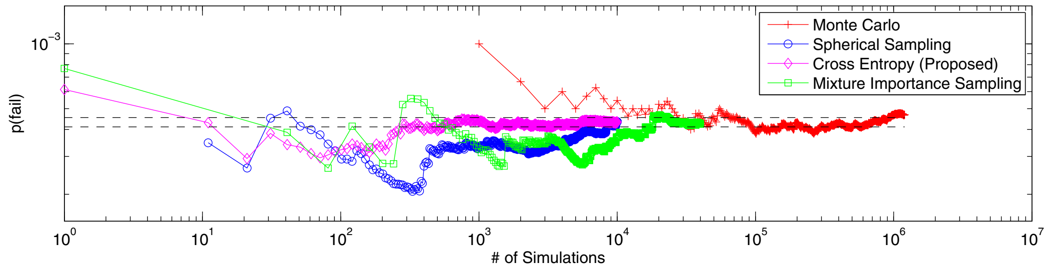
- To achieve a figure of merit of ρ = 0.1 for $V_{DD} = 300mV$, the proposed method requires 561 simulations when compared to 9,771 simulations by SS and 28,500 simulations by MixIS. This leads to computational savings of the order of ∼17X when compared to SS and ∼50X when compared to MixIS. It can be observed that even for 1.2E+6 simulations, MC has not been able to achieve the desired figure of merit and we achieve more than ∼ 2140X speedup over it.
- If we draw a vertical line at 10,000 simulations (i.e., same number of simulations for all the approaches) and the value of figure of merit is compared (Table I) then we can observe that the proposed approach gives a phenomenal figure of merit of ~ 0.02 (98% accuracy with 98% confidence) while SS and MixIS give a very ordinary figure of merit of ~ 0.1 and the value for MC is way below for any comparison. This shows the increased accuracy of the proposed method.

3) Efficiency: Now that the accuracy and the significant computational savings of the proposed approach are established, we proceed to demonstrate its efficiency. In particular, a closer look at the plot of evolution of the figure of merit (see Fig.2(b)) reveal the following interesting observations:

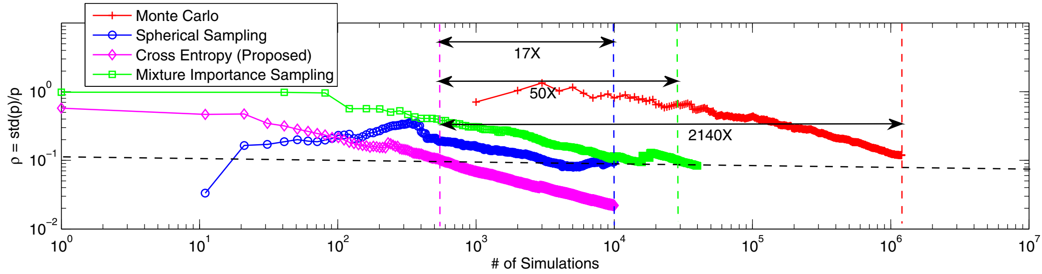
- The *rate of improvement* of figure of merit (slope of the evolution) with increasing number of simulations demonstrate the superiority of the proposed approach. As can be seen in Fig.2(b), the other approaches show a relatively *slow* improvement in the figure of merit whereas the proposed approach has relatively fast improvement which points to faster convergence in the estimates.
- The efficiency of the proposed approach is also demonstrated by the experimental values in Table (II) for other V_{DD} levels (i.e. 275mV and 290mV). In particular, the proposed method shows similar behavior such as fast convergence and accurate probability estimation for different cases. Also, when compared with MC, it achieves ∼87.6X speedup for $V_{DD} = 275mV$, $\sim 454X$ speedup for $V_{DD} = 290mV$ and $\sim 2140X$ speedup for $V_{DD} = 300mV$. There is no drop in performance of the approach for any of the V_{DD} levels and infact the performance betters as we move to extremely small failure probabilities (higher values of V_{DD}) which is a clear indicator of the superior performance of the proposed approach for rare events.

V. Conclusion

In this paper, a highly efficient and general method for quick and accurate estimation of SRAM failure rate was proposed. The proposed methodology, ***Cross Entropy Minimization for SRAM***, adapts from the statistics community, the Information-Theory inspired Importance Sampling based Minimum Cross Entropy approach. It offers a simple analytical solution to calculate the *optimal practical distribution* for Importance Sampling, without making any assumptions to the circuit. As is demonstrated by our experimental results, the proposed method for the same accuracy, achieves runtime speed-up of



(a) Estimation of Failure Probability at $V_{DD} = 300mV$



(b) Figure of Merit at $V_{DD} = 300mV$

Fig. 2: Evolution of Estimation of Failure Probability and Figure of Merit at $V_{DD} = 300mV$

TABLE II: Comparison of Efficiency and Computational Speed up at different V_{DD} levels

		MC	MixIS	SS	Proposed method
V_{DD} = 275mV	p_{fail}	1.82E-3	2.282E-3	2.678E-3	2.33E-3
	Accuracy	90%	90%	90%	90%
	# of runs	1.22E+05 (1X)	9.5E+4 (1.28X)	2.359E+3 (51.7X)	1393 (87.6X)
V_{DD} = 290mV	p_{fail}	7.823E-4	8.531E-4	8.163E-4	8.078E-4
	Accuracy	90%	90%	90%	90%
	# of runs	2.46E+5 (1X)	5.00E+4 (5X)	1.383E+3 (178X)	542 (454X)
V_{DD} = 300mV	p_{fail}	4.675E-4	4.332E-4	4.208E-4	4.225E-4
	Accuracy	88%	90%	90%	90%
	# of runs	1.2E+6 (1X)	2.85E+4 (42X)	9.771E+3 (123X)	561 (2140X)

$17 \sim 50X$ over other state-of the-art techniques. It can be further extended to calculate the timing delay in statistical static timing analysis problems and also to efficiently estimate the parametric yield of high dimensional SRAM circuits with multiple failure regions.

REFERENCES

[1] K. Agarwal and S. Nassif, "Statistical analysis of sram cell stability," in *Proceedings of the 43rd annual Design Automation Conference*, ser. DAC '06, 2006, pp. 57–62.

[2] R. Heald and P. Wang, "Variability in sub-100nm sram designs," in *Computer Aided Design, 2004. ICCAD-2004. IEEE/ACM International Conference on*, 2004, pp. 347–352.

[3] 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 *Proceedings of the 43rd annual Design Automation Conference*, ser. DAC'06, 2006, pp. 69–72.

[4] 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.

[5] 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.

[6] A. Singhee and R. 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, 2007. DATE '07*, 2007, pp. 1–6.

[7] C. Dong and X. Li, "Efficient sram failure rate prediction via gibbs sampling," in *Proceedings of the 43rd annual Design Automation Conference*, ser. DAC'11, 2011.

[8] A. Bayrakci, A. Demir, and S. Tasiran, "Fast monte carlo estimation of timing yield with importance sampling and transistor-level circuit simulation," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 29, no. 9, pp. 1328 –1341, sept. 2010.

[9] H. T. H. O. Kentaro Katayama, Shiho Hagiwara and T. Sato, "Sequential importance sampling for low-probability and high-dimensional sram yield analysis," in *Proceedings of the 2010 IEEE/ACM International Conference on Computer-Aided Design*, ser. ICCAD '10, 2010.

[10] J. Wang, A. Singhee, R. A. Rutenbar, and B. H. Calhoun, "Two fast methods for estimating the minimum standby supply voltage for large srams," *Trans. Comp.-Aided Des. Integ. Cir. Sys.*, vol. 29, pp. 1908–1920, 2010.

[11] R. R.Y. and K. D.P., "The cross-entropy method: a unified approach to combinatorial optimization, monte-carlo simulation and machine learning," *Springer-Verlag*, 2004.

[12] C. T.M. and T. J.A., "Elements of information theory," *John Wiley and Sons*, 1991.

[13] C. Morris, "Natural exponential families with quadratic variance functions," *The Annals of Statistics*, vol. 10, pp. 65–80, 1982.