A Non-Gaussian Adaptive Importance Sampling Method for High-Dimensional and Multi-Failure-Region Yield Analysis

[1]Xiao Shi, [2]Hao Yan, [2]Chuwen Li, [3]Jianli Chen, [2]Longxing Shi, [1]Lei He
[1]Electrical and Computer Engineering Dept., University of California, Los Angeles, CA, USA
[2]Electrical Engineering Dept., Southeast University, China
[3]State Key Lab of ASIC & System, Microelectronics Dept., Fudan University, China
pokemoon2009@g.ucla.edu,yanhao@seu.edu.cn,lichuwens@outlook.com,chenjianli115@126.com,lxshi@seu.edu.cn,lhe@ee.ucla.edu

ABSTRACT

Rare-event yield analysis is challenging for high-dimensional circuit cases. In this paper, we propose a non-Gaussian adaptive importance sampling (NGAIS) method. In order to approximate the failure region in high-dimensional space, we model it as a mixture of von Mises-Fisher distributions. We formulate the parameter estimation problem as a maximum likelihood estimation problem, and then solve with expectation-maximization algorithm. Experiments on bit cell, amplifier and SRAM column circuit validate that the proposed NGAIS method outperforms other state-of-the-art approaches in terms of accuracy and efficiency.

KEYWORDS

Process Variation, Failure Probability, Adaptive Importance Sampling, von Mises-Fisher distribution, Maximum Likelihood Estimation

1 INTRODUCTION

As microelectronic devices shrink to the nano-meter scale, process variation has become a growing concern due to its increasingly significant impact on circuit reliability. Among various integrated circuits, SRAM circuits are highly duplicated with minimum size devices that require extremely small failure probability, which makes it a "rare-event" problem. To ensure the robustness of such circuits, we require a precise estimation of rare circuits failure probability in the pre-silicon phase.

Generally, high-sigma failure probability estimation is achieved by modern stochastic circuit simulation methods. Among these methods, standard Monte Carlo (MC) method is recognized as a gold standard. It repeatedly collects samples and runs transistor-level simulations to determine whether circuits are fail. However, MC does not apply to high-sigma case because we need to perform millions of simulations to capture one single failure event, which is extremely time-consuming.

Prior Work. To replace expensive MC simulations, more efficient approaches have been proposed to collect samples from the likely-to-fail regions, which can be classified into three major categories:

(1) Classification: Methods in this category construct classifiers to filter out of samples that unlikely to fail so they can explore boundaries of failure regions. The failure probability is evaluated by computing hypervolumn of failure regions without extra simulations. To reduce classification error, [1] introduces safety margin to avoid misclassification. More recently, [2, 3] separately design conditional classifier and SVM-based nonlinear classifier to solve multi-failure-region circuits. However, samples required for training classifiers increase sharply when facing high-dimensional circuit cases.

(2) Meta-modeling: Modeling methods are developed in recent works to build circuit evaluators to substitute circuit simulators. The main idea is to map the variation parameters with the circuit metrics. Compared with directly running circuit simulators, evaluating meta-models drastically reduces computational complexity during yield estimation. For example, the approaches in [4, 5] build surrogate models based on Gaussian process regression and radial basis function network, respectively. Moreover, the approach in [6] utilizes low-rank tensor approximation (LRTA) method to train an efficient meta-model. However, the estimation accuracy is highly sensitive to the accuracy of the meta-model, which generally requires a massive number of training samples.

(3) Importance Sampling (IS): Methods in this category try to sample from a shifted distribution that covers likely-to-fail regions. As a classic modification of MC method, IS can improve the estimation accuracy and meanwhile accelerate the convergence. Different IS methods apply various techniques to construct shifted distribution. For example, [7] combines a shifted version of the original distribution and a uniform distribution, [8] shifts the original distribution to the mean of initial failure samples, and [9] constructs multiple mean-shift vectors. Moreover, various strategies have been proposed to modify the conventional static IS sampling distribution. The method in [10] uses multi-start-point sequential quadratic programming to search for optimal shift vectors. The method in [11] utilizes a resampling scheme to sample from the regions with higher importance. However, the drawbacks of these importance sampling methods are that their performance significantly relies on the choice of shift vector, and computational efficiency decreases exponentially with circuit dimensions.

Paper Contributions. In this paper, we propose a novel and efficient non-Gaussian adaptive importance sampling to tackle the challenging high-dimensional yield analysis problem. The specific contributions include:

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
ICCAD '20, November 2–5, 2020, Virtual Event, USA
© 2020 Association for Computing Machinery.
ACM ISBN 978-1-4503-8026-3/20/11…$15.00
https://doi.org/10.1145/3400302.3415633

- An adaptive scheme to explore high-dimensional space and search for failure regions. The sampling procedure includes a series of incremental sampling iterations and successively updates partial IS estimators. The adaptation of intermediate sampling distribution is driven by maximum likelihood estimation to move toward the target failure region, which can provide a more accurate estimation.
- An innovative von Mises-Fisher (vMF) mixture distribution as IS distribution. It samples from a hyperspherical surface with excellent flexibility in directions and dispersion, which corresponds to the spatial feature of multivariate Gaussian distribution in high dimension. Moreover, the multi-modal vMF mixture can approximate multiple disjoint failure regions in parallel, which is critical for the circuit with various failure mechanisms.
- An efficient and effective solution to parameter estimation. We formulate it into a maximum likelihood estimation problem with hidden variables and implement an expectation maximization algorithm to solve for the parameters.

2 PRELIMINARIES

2.1 Rare Event Analysis

Let variable X denotes d-dimensional circuit process variation and its joint distribution can be defined as a multivariate probability density function (PDF) $f(X)$. In general, the variable X is modeled as a multivariate normal distribution by manufacture. Let Y represent the circuit performance of interest, such as memory read/write time, amplifier gain, etc. This performance needs to meet the predefined threshold to classify it as the "PASS" sample by expensive transistor-level circuit simulation.

In order to express this relationship mathematically, we introduce an indicator function $I(X)$ to identify pass/fail of Y:

$$I(X) = \begin{cases} 0, & if \quad Y \notin S \\ 1, & if \quad Y \in S \end{cases} \tag{1}$$

where S denotes the failure region which includes all the failed samples. Then the failure probability can be evaluated as follows:

$$P_{fail} = P(Y \in S) = \int I(X) \cdot f(X) dX \tag{2}$$

However, the integral in Equation (2) is intractable because $I(X)$ is unavailable in analytical form. Traditional Monte Carlo (MC) method enumerates a sample set $\{x_i\}_{i=1}^{N}$ according to $f(X)$ and evaluate their indicator function values to generate an unbiased estimate of $\hat{P}_{fail}$:

$$\hat{P}_{fail} = \hat{P}(Y \in S) = \frac{1}{N} \sum_{i=1}^{N} I(x_i) \xrightarrow{N \to +\infty} P(Y \in S) \tag{3}$$

2.2 Importance Sampling

For rare-event analysis, such as circuit failure event, the traditional MC method collects enough failed samples through simulating millions of samples, which can calculate a credible failure probability estimation. As shown in Figure 1, in order to reduce the simulation numbers dramatically, IS has been designed to draw samples from a specially proposed distribution $g(X)$ that tilts towards failure region S, which can generate numerous failed samples.

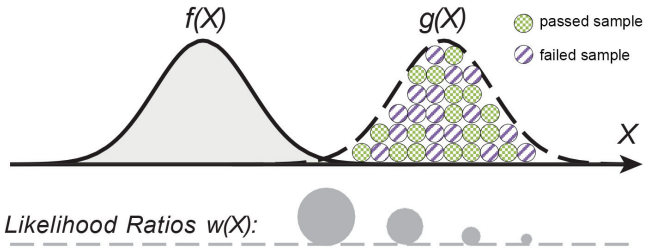

Figure 1: Mean-shift importance sampling

Then the failure probability can be computed analytically as follows:

$$P_{fail} = P(Y \in S) = \int I(X) \cdot \frac{f(X)}{g(X)} \cdot g(X) dX \tag{4}$$

$$= \int I(X) \cdot w(X) \cdot g(X) dX = E_{g(X)}[I(X) \cdot w(X)] \tag{5}$$

where $w(X)$, usually called importance weight, represents the likelihood ratio between original PDF $f(X)$ and the proposed PDF $g(X)$. According to the Law of Large Numbers, the unbiased IS estimator $\hat{P}_{IS,fail}$ can be calculated straightforwardly based on samples $\{x_j\}_{j=1}^{M}$ from $g(X)$.

$$\hat{P}_{IS,fail} = \hat{P}(Y \in S) = \frac{1}{M} \sum_{i=1}^{M} w(x_j) I(x_j) \xrightarrow{M \to +\infty} P(Y \in S) \tag{6}$$

Notice that the sample size in Equation (6) is much smaller than the MC samples size in Equation (3) because the failure in $g(X)$ is not a "rare event". Theoretically, the failure event distribution $\pi(X)$ is just the optimal sampling distribution $g^{opt}(X)$:

$$g^{opt}(X) = \frac{I(X) \cdot f(X)}{P_{fail}} \tag{7}$$

However, as the function $I(X)$ is unknown in analytical expression and P_{fail} is indeed the desired failure probability, $g^{opt}(X)$ cannot be evaluated with Equation (7) directly. Therefore, most existing methods contribute various strategies to approximate the failure event distribution. For example, the method [12] shifts the sample mean toward either pass/fail boundary, the method [8] sets the centroid of presampling failure points as the sample mean, the method [9] sets multiple centroids by clustering failure points as the sample means.

In fact, there are two major drawbacks with these IS based methods. First, these methods are extremely sensitive to the presampling stage because the PDF $g(X)$ is directly constructed based on the failed samples from presampling. Next, the static IS methods lack of flexibility to explore the entire parametric space. These methods become less effective in high-dimensional space, especially when the failure region has a complicated boundary.

3 NON-GAUSSIAN ADAPTIVE IMPORTANCE SAMPLING ALGORITHM

3.1 Algorithm Description

The main steps of our proposed NGAIS algorithm are summarized in Algorithm 1. In the initialization step, we start with a failure samples set $\{x_i^{(0)}\}_{i=1}^N$ collected with LHS method, which is a space-filling sampling method in a unified grid formulation. At each iteration t, we approximate the failure region with a mixture of vMF distributions, followed by an EM framework to solve the parameters. Next we generate N new samples $\{x_i^{(t)}\}_{i=1}^N$ from this vMF mixture and compute their importance weights $\{w_{i,t}\}_{i=1}^N$. At the end of each iteration, we update the partial IS estimator by averaging all importance weights $\{w_{i,t}\}_{i=1}^{tN}$. This iterative process continues until our estimation converges to a certain confidence interval.

We first show that our partial IS estimator can always converge to correct failure probability when sufficient large sample size is allowed. The unbiasedness of our NGAIS estimator can be validated by its expected value:

$$E[\hat{P}_{fail}] = E[\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N} w_{i,t}] \tag{8}$$

$$= \frac{1}{N}\sum_{i=1}^{N} E[\frac{f(X)I(X)}{g(X)}] \tag{9}$$

$$= \frac{1}{N}\sum_{i=1}^{N}\int \frac{f(X)I(X)}{g(X)} g(X)dx = P_{fail} \tag{10}$$

Therefore, our NGAIS algorithm can embrace the flexibility of sampling distribution adaptation, but still provide very consistent estimation.

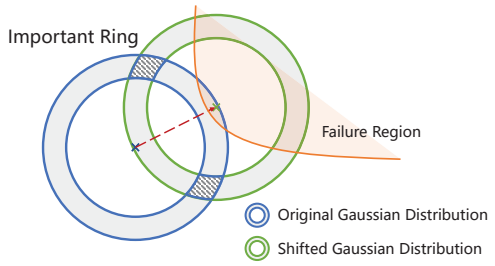

Figure 2: Geometric illustration of mean-shift IS method. The blue and green important rings belong to original Gaussian distribution $f(X)$ and shifted distribution, respectively. The orange shaded region represents the failure region. The intersection between important rings can hardly cover the failure region.

However, the major challenge of the Gaussian Mixture based IS method is how to effectively approximate the failure region distribution to generate more "important" failure samples in high dimension. The rationale behind this problem is that Euclidean norm is distorted in high dimension. Considering a multivariate Gaussian distribution $f(X)$ in d-dimensional parametric space, the samples generated from $f(X)$ follow that the Euclidean distance between the sample and the Gaussian mean displays a χ-distribution with d degrees of freedom: $r \sim \chi_d$. In the extreme case when the dimension d increases to infinity, the χ-distribution $\chi_d \approx N(\sqrt{d}, 1/2)$, which means it can be approximated as a hyperspherical surface at $\sqrt{d}$.

Algorithm 1: NGAIS Algorithm

Initialization:

Set iteration index $t = 0$ and generate the initial failed samples set $\{x_i^{(t)}\}_{i=1}^N$ through LHS sampling.

repeat

Update iteration index $t = t + 1$.

1. Intermediate Sampling Distribution:

Construct a parameterized vMF mixture as intermedediate sampling distribution

$$g(X) = \sum_{k=1}^{K} \alpha_k v_k(X|\mu_k, \kappa_k)$$

2. Maximum Likelihood Estimation:

Formulate the log-likelihood function with failed samples set $\{x_i^{(t)}\}_{i=1}^N$:

$$\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\kappa}) = \sum_{i=1}^{N} ln(\alpha_{z_i} v_{z_i}(x_i^{(t)}|\mu_{z_i}, \kappa_{z_i}))$$

3. Expectation-Maximization Algorithm:

for $m = 1, 2, \cdots, M$ **do**

(a) E-step: Calculate the expectation of the likelihood function $E[\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\kappa})]$.

(b) M-step: Maximize the expectation to solve the parameters: $argmax_{\boldsymbol{\alpha},\boldsymbol{\mu},\boldsymbol{\kappa}} E[\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\kappa})]$

end

4. Samples Propagation:

Generate a new set of N samples $\{x_i^{(t)}\}_{i=1}^N$ in the parametric space:

$$x_i^{(t)} \sim g(X) \qquad i = 1, 2, \cdots, N$$

5. Weighting:

Compute importance weights:

$$w_{i,t} = \frac{\pi(x_i^{(t)})}{g(x_i^{(t)})} = \frac{f(x_i^{(t)})I(x_i^{(t)})}{g(x_i^{(t)})} \quad i = 1, \ldots, N$$

6. Estimation:

Update the unbiased estimator:

$$\hat{P}_{fail,t} = \frac{1}{tN}\sum_{j=1}^{t}\sum_{i=1}^{N} w_{i,j} \quad i = 1, \ldots, N$$

until *Relative standard deviation (FOM):* $\rho = \frac{\sqrt{\sigma^2_{\hat{P}_{fail}}}}{\hat{P}_{fail}}$;

As shown in Figure 2, with increasing dimension, the mass of a multivariate Gaussian distribution is not near the mean, but instead forms an "important ring", which is depicted in blue. In order to

illustrate the conventional mean-shift IS method, we notice that the intersection between original Gaussian distribution and the shifted one is tiny. In other words, in order to shift Gaussian distribution $g(X)$ to cover the failure region located inside the "important ring", $g(X)$ would have very small variance, which could lead to numerical issues.

3.2 Mixture vMF Distribution for High-Dimensional Samples

In this section, we explain how we utilize a mixture of vMF distributions to model the failure region in high dimension. vMF distribution is a model for directional samples distributed uni-modally with rotational symmetry, which is defined on the unit hypersphere. The intuition behind vMF distribution is analogy to the multivariate Gaussian distribution in spherical coordinates. The mathematical expression of a single-modal vMF distribution is written as:

$$v(X|\mu,\kappa) = c_d(\kappa)e^{\kappa\mu^T X} \tag{11}$$

Here the unit mean direction vector μ is analogous to the mean vector of multivariate Gaussian distribution, but in radial direction. The concentration parameter κ is a measure of the degree of directional dispersion, which is analogous to the variance of multivariate Gaussian distribution. In the extreme case, when $\kappa = 0$, vMF distribution will reduce to uniform distribution on $(d-1)$-dimensional unit sphere; when $\kappa = \infty$, it will reduce to a single point. Further, c_d is a normalizing constant given by:

$$c_d(\kappa) = \frac{\kappa^{d/2-1}}{(2\pi)^{d/2} I_{d/2-1}(\kappa)} \tag{12}$$

where $I_{d/2-1}(\cdot)$ is the $(d/2-1)^{th}$ order modified Bessel function of the first kind.

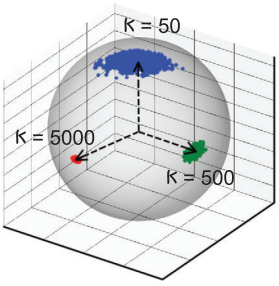

Figure 3: An example vMF mixture distribution in three-dimensional space. The dashed arrows denote the mean direction vectors and concentration parameters are also labeled separately.

The geometric insight of vMF distribution is to generate samples on the surface of unit sphere. As we have explained in Section 3.2, in d-dimensional parametric space, the samples are distributed inside an important ring whose radius is $[\sqrt{d}-\epsilon, \sqrt{d}+\epsilon]$. And the thickness of this important ring 2ϵ is extremely thin when the dimension d is high. We can see that the probabilistic density of samples is desirable to be modeled as vMF distribution. It is because tuning the parameters only changes the direction and concentration of samples, while preserving the same radius. In the extreme case when $d \to \infty$, the important ring will reduce to a hyperspherical surface with radius $\sqrt{d}$, which exactly forms a vMF distribution.

In practical, high-dimensional circuits usually have multiple failure regions induced by various failure mechanisms. Adopting the idea of mixture Gaussian, it is straightforward to extend the single-modal vMF distribution to a mixture by taking the weighted sum of multiple vMF distributions, as illustrated in Figure 3. The mathematical expression is given by:

$$g(X) = \sum_{k=1}^{K} \alpha_k v_k(X|\mu_k,\kappa_k) \tag{13}$$

where K is the number of vMF distributions and α_k is the corresponding normalized weight function. Here the maximal mixture index K is a user-defined parameter that can be tuned. It exhibits a trade-off between model complexity and approximation accuracy. In this work, we set $K = \sqrt{N}$, where N is the current sample size. This is extensively used in the Statistics community [9]. However, given the current sample set, solving for the parameters in Equation (13) is non-trivial. We will describe our solver in the next section.

3.3 Parameter Estimation of vMF Mixture Density

At each iteration, our target is to estimate the parameters of the vMF distributions from a given sample set. We first formulate the mixture-density parameter estimation problem into maximum likelihood estimation. Then we implement an EM framework to solve this problem.

3.3.1 Maximum Likelihood Estimation Problem Formulation. In order to sample from the vMF mixture defined in Equation (13), we assume the k-th vMF $v_k(X|\mu_k,\kappa_k)$ is chosen with probability α_k. Let $\mathcal{X} = \{x_1, x_2, ..., x_N\}$ be a sample set collected from this vMF mixture independently, and $\mathcal{Z} = \{z_1, z_2, ..., z_N\}$ be the corresponding set of hidden variables whose element z_i determines certain sample x_i is drawn from mixture $v_{z_i}(x_i|\mu_{z_i},\kappa_{z_i})$. Thus the log-likelihood of the sample set $\mathcal{X}$ is given by

$$\mathcal{L}(\boldsymbol{\alpha},\boldsymbol{\mu},\boldsymbol{\kappa}) = \sum_{i=1}^{N} ln(\alpha_{z_i} v_{z_i}(x_i|\mu_{z_i},\kappa_{z_i})) \tag{14}$$

However, Equation (14) is infeasible in analytical form because the hidden variable set $\mathcal{Z}$ is unknown.

3.3.2 Solution with Expectation-Maximization Algorithm. In this section, we implement an EM algorithm to solve for the maximum likelihood estimation. EM algorithm iterates between two major steps: an M-step for parameter estimation and an E-step for distribution estimation.

M-step: In this step, we first derive the expectation of log-likelihood function $\mathcal{L}$ into the form of conditional probability of

hidden variable $p(k|x_i, \boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\kappa})$:

$$\begin{aligned} & E[\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\kappa})] \\ &= \sum_{i=1}^{N} E_{p(z_i|x_i,\boldsymbol{\alpha},\boldsymbol{\mu},\boldsymbol{\kappa})}[\ln(\alpha_{z_i} v_{z_i}(x_i|\mu_{z_i}, \kappa_{z_i}))] \\ &= \sum_{i=1}^{N}\sum_{k=1}^{K} \ln(\alpha_k v_k(x_i|\mu_k, \kappa_{z_i}))p(k|x_i, \boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\kappa}) \end{aligned} \tag{15}$$

Then we re-estimate $\{\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\kappa}\}$ to maximize the above expectation. We introduce the Lagrangian multiplier λ and take partial derivatives of the Lagrangian Objective function w.r.t. each α_k:

$$\begin{aligned} & \frac{\partial}{\partial \alpha_k}[\sum_{i=1}^{N}\sum_{k=1}^{K}(\ln \alpha_k)p(k|x_i, \boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\kappa}) + \lambda(\sum_{k=1}^{K}\alpha_k - 1)] = 0 \\ & \Longrightarrow \sum_{i=1}^{N} p(k|x_i, \boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\kappa}) = -\lambda\alpha_k \end{aligned} \tag{16}$$

We sum both sides of Equation (16) over all k to get $\lambda = -N$ and simplify Equation (16) into:

$$\alpha_k = \frac{1}{N}\sum_{i=1}^{N} p(k|x_i, \boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\kappa}) \tag{17}$$

Similarly, we can derive $\boldsymbol{\mu}, \boldsymbol{\kappa}$ in the same form:

$$\mu_k = \frac{\sum_{i=1}^{N} x_i p(k|x_i, \boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\kappa})}{||\sum_{i=1}^{N} x_i p(k|x_i, \boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\kappa})||} \tag{18}$$

$$\frac{I_{d/2}(\kappa)}{I_{d/2-1}(\kappa)} = \frac{||\sum_{i=1}^{N} x_i p(k|x_i, \boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\kappa})||}{\sum_{i=1}^{N} p(k|x_i, \boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\kappa})} \tag{19}$$

E-step: In this step, we update the conditional distribution of the hidden variables $p(k|x_i, \boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\kappa})$ based on the parameter we have solved. The hidden variables describe the distribution assignment from the vMF mixture. For each iteration of EM algorithm, we approximate this distribution assignment as a probabilistic assignment of the samples:

$$p(k|x_i, \boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\kappa}) = \frac{\alpha_k v_k(x_i|k, \boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\kappa})}{\sum_{l=1}^{K} \alpha_l v_l(x_i|l, \boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\kappa})}, \tag{20}$$

4 EXPERIMENTAL RESULTS

In this section, our proposed NGAIS algorithm is first tested on a typical 6T SRAM bit cell with 18 variables. For analog yield analysis, we then validate our NGAIS on a two-stage operational transimpedance amplifier (OTA) with 126 variables. More realistically, we verify our method on an SRAM column circuit with 597 variables, which is a high-dimensional case. We implement MC as ground truth for accuracy comparison. To show the efficiency of NGAIS, we also implement several state-of-the-art approaches including Hyperspherical Clustering and Sampling (HSCS) [9], Adaptive Importance Sampling (AIS) [11] and Multiple Failure Region Importance Sampling (MFRIS) [10]. The SPICE model is SMIC 40nm transistor model. All the experiments are performed on Linux server with Intel Xeon X5675 CPU @3.07 GHz and 94 GB RAM.

4.1 Experiments on 6T SRAM Bit Cell

The schematic of a typical 6T SRAM bit cell is shown in Figure 4. Four transistors form two cross-coupled inverters and use two steady states (either '0' or'1') to store data in the cell. The other two transistors work as switches to control access to the storage cell during read and write operations. Without loss of generality, we consider SRAM reading failure in our experiment. It occurs when the voltage difference between BL and BLB is too small to be captured by sense amplifier in a pre-decided time period. We evaluate different methods(MC, HSCS, MFRIS, AIS, proposed) to compare accuracy and efficiency.

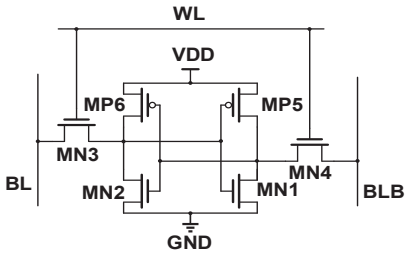

Figure 4: The schematic of typical 6T SRAM cell

4.1.1 ***Accuracy Comparison.*** To evaluate the accuracy of different methods, we utilize Figure of Merit (FOM) ρ to characterize the convergence and confidence interval of estimation, which is defined as:

$$\rho = \frac{\sqrt{\sigma^2_{\hat{P}_{fail}}}}{\hat{P}_{fail}} \tag{21}$$

where $\hat{P}_{fail}$ is the estimation of P_{fail} and $\sigma_{\hat{P}_{fail}}$ denotes its standard deviation. With this definition, we can declare one estimation has $(1-\epsilon)\times 100\%$ accuracy with $(1-\delta)\times 100\%$ confidence when $\rho < \epsilon\sqrt{log(1/\delta)}$. In our experiments, we use $\rho < 0.1$ to indicate the estimation reaches a steady state with 90% confidence interval. ρ has been extensively used in the literature [9, 11].

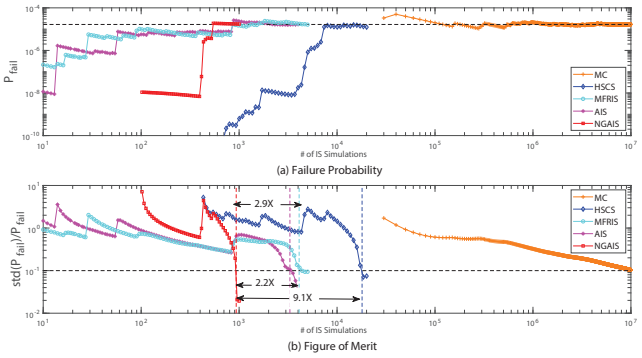

Figure 5: Evolution comparison of failure prob. and FOM on 18-dimensional bit cell

Figure 5(a) demonstrates the evolution of failure probability estimation. We first notice that the failure rate estimations of HSCS,

MFRIS, AIS and proposed NGAIS all converge to ground-truth MC value when sufficient simulations are allowed. The stable estimation value is reached when the FOM is smaller than 0.1, which is denoted by the dashed line in Figure 5(b). The numerical comparison is illustrated in Table 1. The ground truth MC result is 1.66e-5 (4.26σ). Among these methods, NGAIS is the most accurate one with only 1.2% relative error.

Table 1: Accuracy and efficiency comparison on 18-dimensional SRAM bit cell

	MC	HSCS	MFRIS	AIS	NGAIS
Failure prob.	1.66e-5	1.77e-5	1.61e-5	1.62e-5	1.64e-5
Relative error	golden	6.6%	3%	2.4%	1.2%
Presampling # sim.	0	4000	3000	2000	1500
IS # sim.	1e7	18002	4100	3300	934
Total # sim.	1e7	22002	7100	5300	2434
Speedup	1X	454X	1408X	1887X	4108X

4.1.2 ***Efficiency Comparison.*** Figure 5(b) shows the efficiency comparison of MC, HSCS, MFRIS, AIS, and NGAIS. For this low-dimensional circuit case, HSCS and MFRIS are less efficient because they spend more simulations to set up the initial sampling distribution. Instead, the adaptive sampling strategy adopted by AIS and NGAIS algorithm gradually evolves to search for failure regions, which is much more efficient than the static sampling distribution. As detailed in Table 1, HSCS, MFRIS, and AIS require 22002, 7100, and 5300 samples in total to converge to golden reference. In comparison, NGAIS obtains higher accuracy with only 2434 simulations, which exhibits 9.1X, 2.9X, and 2.2X speedup over existing methods, respectively.

4.2 Experiments on Two-stage Amplifier

Next, we verify that the proposed method is capable of handling problems on an analog circuit with multiple specification requirements. A simplified schematic of the two-stage operational transimpedance amplifier (OTA) consisting of master-slave structure for low supply voltage application is presented in Figure 6. The slave stage consists of the tail current transistor (MP5), the differential pair (MP1 and MP2) and the current-mirror load (MN1 and MN2). The master stage replicates the tail current source and the transconductance transistor of the slave stage (i.e. MP3, MP4, MP6 and MN3 are copies of MP1, MP2, MP5 and MN1). MP5 and MP6 operate in the linear region to save voltage margin.

The estimation of P_{fail} for OTA circuit is more troublesome because it has various failure mechanisms, which lead to multiple failure regions. In our experimental setting, we consider various performance specifications, including voltage gain margin, gain bandwidth, phase margin, and 3dB bandwidth. There are in total 126 variation parameters in this case.

4.2.1 ***Accuracy Comparison.*** In order to validate the accuracy of our algorithm, we plot the evolution of failure probability estimation and FOM vs. number of IS simulations in Figure 7. To generate ground truth, MC takes 4.59e5 simulations to generate confident estimation of P_{fail} at 1.5e-3 (3.19σ). For this multi-performance

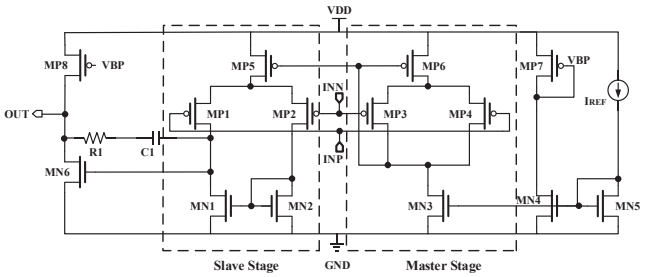

Figure 6: The schematic of two-stage amplifier circuit

circuit case, we first notice that all four different methods are able to tackle circuits with multiple failure regions. It is also worth noting that MFRIS and NGAIS are more accurate ones with less than 10% relative error, while HSCS and AIS have 16% and 10.6% relative error, respectively. It is attributed to the better capability of fitting multi-modal failure region distributions. To be specific, MFRIS implements a multi-start-point sequential quadratic programming framework to perform local optimizations and search for a set of optimal shift vectors. Our NGAIS develops an adaptive scheme to updates intermediate sampling distributions by successively solving the maximum likelihood estimation problem. In comparison, HSCS and AIS directly use Gaussian mixture distribution to approximate the failure region distribution, which induces larger relative error.

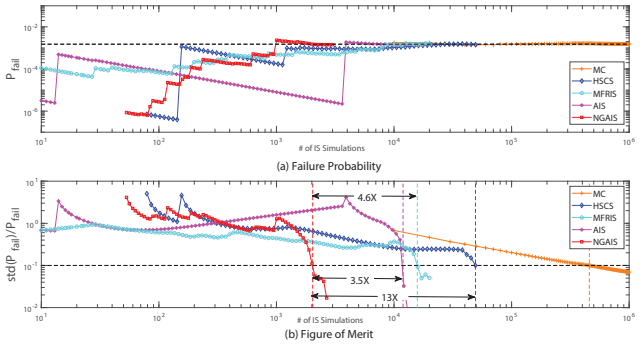

Figure 7: Evolution comparison of failure prob. and FOM on 126-dimensional two-stage amplifier circuit

4.2.2 ***Efficiency Comparison.*** Then we compare different methods in terms of efficiency. Evolution of P_{fail} convergence and FOM evaluation are plotted in Figure 7, the following observations can be made:

- First, we notice that the P_{fail} curve of our NGAIS method changes more frequently towards the golden MC result. This flexibility is attributed to our carefully designed vMF distribution. Compared with Gaussian mixture, it is more suitable for modeling failure region distribution with multiple peaks. This feature accelerates our estimation procedure and exhibits better convergence property.
- Second, the FOM curve of our NGAIS method decreases faster before asymptotically reaching 90% confidence. It is

because NGAIS performs a global approximation by successively solving for maximum likelihood problems. It can provide very consistent estimation as the sample set grows.

Table 2: Accuracy and efficiency comparison on two-stage amplifier circuit

	MC	HSCS	MFRIS	AIS	NGAIS
Failure prob.	1.5e-3	1.26e-3	1.6e-3	1.34e-3	1.42e-3
Relative error	golden	16%	6.7%	10.6%	5.3%
Presampling # sim.	0	7000	4000	3100	2275
IS # sim.	458700	49290	15776	11930	2035
Total # sim.	458700	56290	19776	15030	4310
Speedup	1X	8X	23X	30X	106X

4.3 Experiments on SRAM Column Circuit

Figure 8 shows a simplified schematic of SRAM column circuit. It consists of 32 bit cells and a sense amplifier with totally 597 variation parameters, which is a high-dimensional problem. Compared with the single bit cell low-dimensional set up in the first part, we consider a more realistic scenario with the impact of peripheral circuits and other idle bits, thus generate a more accurate failure probability estimation. We configure the worst case of read operation in Figure 8, in which accessing bit $CELL < 0 >$ stores "0" when other idle bits all store "1". In this case, the leakage current increases read access time and impedes successful read.

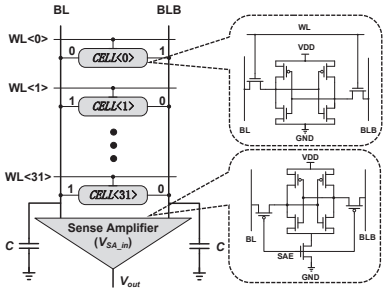

Figure 8: The schematic of 597-dimensional SRAM column with peripheral circuits

4.3.1 ***Accuracy and Efficiency Comparison.*** Figure 9(a) compares the evolution of failure probability estimation between different methods. For this high-dimensional circuit case, we first notice that only MFRIS and our proposed NGAIS can converge to ground-truth MC value. HSCS and AIS converge to biased results, whose relative error is order of magnitude. It is because the static and deterministic Gaussian mixture used in HSCS cannot effectively cover failure regions in high-dimensional parametric space. And the resampling procedure in AIS tends to neglect less important failure regions in high dimension, which leads to a smaller P_{fail}. To tackle the dimensionality issue, MFRIS implements an adaptive model training framework that consists of multiple local modeling steps, and uses a surrogate model to reduce the computational cost

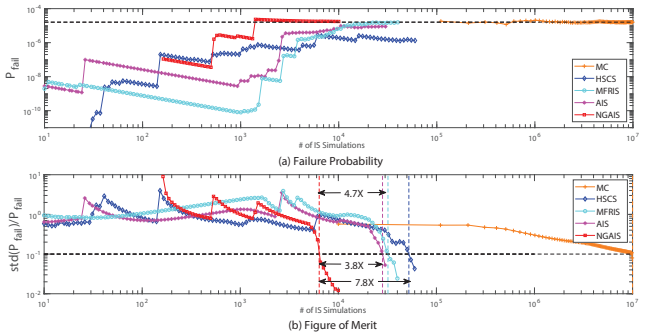

Figure 9: Evolution comparison of failure prob. and FOM on 597-dimensional SRAM column with peripheral circuits

of IS. Our NGAIS method handles this high-dimensional problem from a different perspective. We take advantage of the geometric aggregation of samples on the "important ring", and approximate their distribution as vMF distribution. It gives us the highest accuracy among these four methods, which is only 4.5% relative error.

Table 3 illustrates the efficiency comparison between different methods. The ground-truth MC estimation is 1.6e-5 (4.34σ). Our NGAIS method is capable of achieving 90% confidence estimation by in total 8913 simulations. In comparison, MFRIS is much slower because it is extremely expensive to train a surrogate model in high-dimensional space. As a conclusion, NGAIS method can achieve 1122X speedup over MC, and 4.7X over MFRIS.

Table 3: Accuracy and efficiency comparison on 597-dimensional SRAM column with peripheral circuits

	MC	HSCS	MFRIS	AIS	NGAIS
Failure prob.	1.6e-5	1.47e-6	1.469e-5	8.93e-6	1.672e-5
Relative error	golden	error	8.1%	error	4.5%
Presampling # sim.	0	18000	10000	6000	2500
IS # sim.	1e7	52377	32057	28224	6413
Total # sim.	1e7	70377	42057	34224	8913
Speedup	1X	142X	238X	292X	1122X

5 CONCLUSION

In this paper, we propose a non-Gaussian adaptive importance sampling algorithm to efficiently estimate extremely small failure probability of high-dimensional and multi-failure-region circuits. We first model our sampling distribution as a vMF mixture density to approximate the target failure event distribution. Next we formulate the parameter estimation problem into maximum likelihood estimation, which is solved by EM method. The experimental results show that NGAIS can provide extremely high accuracy and efficiency. For SRAM bit cell with 18 variables, NGAIS has 4108X speedup over MC, and 2-9X speedup over other state-of-the-art methods. For the two-stage amplifier circuit, NGAIS achieves the highest accuracy with 3.5-13X acceleration. For the 597-dimensional SRAM column with peripheral circuits, NGAIS is 1122X faster than MC method, while other approaches either converge to wrong results or require much more simulations to converge.

REFERENCES

[1] Amith Singhee and Rob A Rutenbar. Statistical blockade: a novel method for very fast monte carlo simulation of rare circuit events, and its application. In *2007 Design, Automation & Test in Europe Conference & Exhibition*, pages 1–6. IEEE, 2007.

[2] Amith Singhee, Jiajing Wang, Benton H Calhoun, and Rob A Rutenbar. Recursive statistical blockade: An enhanced technique for rare event simulation with application to sram circuit design. In *VLSI Design, 2008. VLSID 2008. 21st International Conference on*, pages 131–136. IEEE, 2008.

[3] Wei Wu, Wenyao Xu, Rahul Krishnan, Yen-Lung Chen, and Lei He. Rescope: High-dimensional statistical circuit simulation towards full failure region coverage. In *Proceedings of the 51st Annual Design Automation Conference*, pages 1–6. ACM, 2014.

[4] Jian Yao, Zuochang Ye, and Yan Wang. An efficient sram yield analysis and optimization method with adaptive online surrogate modeling. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 23(7):1245–1253, 2014.

[5] Mengshuo Wang, Wenlong Lv, Fan Yang, Changhao Yan, Wei Cai, Dian Zhou, and Xuan Zeng. Efficient yield optimization for analog and sram circuits via gaussian process regression and adaptive yield estimation. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 37(10):1929–1942, 2018.

[6] Xiao Shi, Hao Yan, Qiancun Huang, Jiajia Zhang, Longxing Shi, and Lei He. Meta-model based high-dimensional yield analysis using low-rank tensor approximation. In *Proceedings of the 56th Annual Design Automation Conference 2019*, pages 1–6, 2019.

[7] Rouwaida Kanj, Rajiv Joshi, and Sani Nassif. Mixture importance sampling and its application to the analysis of sram designs in the presence of rare failure events. In *Design Automation Conference, 2006 43rd ACM/IEEE*, pages 69–72. IEEE, 2006.

[8] Wei Wu, Fang Gong, Gengsheng Chen, and Lei He. A fast and provably bounded failure analysis of memory circuits in high dimensions. In *2014 19th Asia and South Pacific Design Automation Conference (ASP-DAC)*, pages 424–429. IEEE, 2014.

[9] Wei Wu, Srinivas Bodapati, and Lei He. Hyperspherical clustering and sampling for rare event analysis with multiple failure region coverage. In *on International Symposium on Physical Design*, pages 153–160, 2016.

[10] Mengshuo Wang, Changhao Yan, Xin Li, Dian Zhou, and Xuan Zeng. High-dimensional and multiple-failure-region importance sampling for sram yield analysis. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 25(3):806–819, 2017.

[11] Xiao Shi, Jun Yang, Fengyuan Liu, and Lei He. A fast and robust failure analysis of memory circuits using adaptive importance sampling method. In *2018 55th ACM/ESDA/IEEE Design Automation Conference (DAC)*, pages 1–6. IEEE, 2018.

[12] Lara Dolecek, Masood Qazi, Devavrat Shah, and Anantha Chandrakasan. Breaking the simulation barrier: Sram evaluation through norm minimization. In *Proceedings of the 2008 IEEE/ACM International Conference on Computer-Aided Design*, pages 322–329. IEEE Press, 2008.