Fast and Accurate High-Sigma Failure Rate Estimation through Extended Bayesian Optimized Importance Sampling

Michael Hefenbrock, Dennis D. Weller, Michael Beigl and Mehdi B. Tahoori
Karlsruhe Institute of Technology, Karlsruhe, Germany
{michael.hefenbrock, dennis.weller, michael.beigl, mehdi.tahoori}@kit.edu

***Abstract*—Due to the aggressive technology downscaling, process variations are becoming pre-dominent, causing performance fluctuations and impacting the chip yield. Therefore, individual circuit components have to be designed with very small failure rates to guarantee functional correctness and robust operation. The assessment of high-sigma failure rates however cannot be achieved with conventional Monte Carlo (MC) methods due to the huge amount of required time-consuming circuit simulations. To this end, Importance Sampling (IS) methods were proposed to solve the otherwise intractable failure rate estimation problem by focusing on high-probable failure regions. However, the failure rate could largely be underestimated while the computational effort for deriving them is high. In this paper, we propose an eXtended Bayesian Optimized IS (XBOIS) method, which addresses the aforementioned shortcomings by deployment of an accurate surrogate model (e.g. delay) of the circuit around the failure region. The number of costly circuit simulations is therefore minimized and estimation accuracy is substantially improved by efficient exploration of the variation space. As especially memory elements occupy a large amount of on-chip resources, we evaluate our approach on SRAM cell failure rate estimation. Results show a speedup of about 16x as well as a two orders of magnitude higher failure rate estimation accuracy compared to the best state-of-the-art techniques.**

I. Introduction

In order to guarantee high functional densities, the number of transistors per chip are still growing in size. At the same time, process variations become more dominant and have a significant impact on circuit performance, robustness and yield [1]. Especially memory components such as SRAM cells, which occupy a large portion of the chip, suffer from process variations leading to reduced chip yield [2]. For this reason, individual SRAM cells require very low failure rates [3], [4] which has to be considered during design phase to allow for safety margins.

In this regard, accurate (SPICE) circuit simulators are important tools utilized to derive all the relevant circuit characteristics required for the yield assessment. However, as those circuit simulations are inherently very time-consuming, one problem arises when it comes to estimating failure rates associated with very rare events. Conventional Monte Carlo (MC) methods, which are based on a statistical model of the circuit, become infeasible in these cases due to the high number of simulations required to achieve reasonable convergence rates. As a resort, Importance Sampling (IS) techniques were proposed in the literature [5] circumventing the problem of MC by sampling from regions with higher failure rate and therefore reducing the required number of circuit simulations for convergence. Nevertheless, finding the correct position (shift vector) of an IS sampling distribution is crucial, as unsuitable shifts can lead to underestimations of the failure rate. Several methods have therefore been proposed in order to find good shift vectors.

In [6], a gradient-based shift vector extraction technique was proposed - called Gradient Importance Sampling (GIS) - which finds failures at the pass-/fail-border using local optimization on the simulator output (e.g. delay) with respect to variation space parameters (e.g. transistors). Although this technique was superior to existing IS approaches, higher probable failure regions were overlooked due to the utilized local optimization routine and the non-convex simulator output. In [7], Bayesian Optimized Importance Sampling (BOIS) was introduced, which utilized Bayesian Optimization in combination with a local optimization procedure to find more favourable shift vectors. Although the failure rate estimation accuracy was substantially improved, the IS runtime was similar. In [8], a framework is presented in which a regression model is build based on uniformly sampled circuit simulations that are afterwards used to predict the delays of IS samples. Although this reduces the runtime of IS substantially, non-guided uniform sampling can however lead to bad approximations of the actual region of interest around the critical delay.

In this work, we address the aforementioned shortcomings by proposing ”eXtended Bayesian Optimized Importance Sampling” (XBOIS), combining the strengths of [7] and [8]. In contrast to the uniform sampling of [8], our method explores the variation space efficiently through minimizing its uncertainty about the failure region (i.e. pass-/fail-border) using Bayesian Optimization. Furthermore, we find the optimal shift vector for IS utilizing a local optimization on a surrogate model of the simulation output (as in BOIS [7]). Due to the quality of the model, *we are able to avoid simulations for IS entirely by replacing them with cheaper model evaluations achieving substantial speedups of up to 16x.*

As especially memory elements such as SRAM cells suffer from process variation-induced performance fluctuations [2], the evaluation of the proposed method is carried out on 6T-SRAM cell arrays for several scenarios, ranging from low-sigma (rare events) to high-sigma (very rare events) and from low-dimensional ($<$10 transistors under variation) to high-dimensional ($\geq$10 transistors under variation) problems.

The rest of the paper is structured as follows: in Section II, we review IS techniques for solving the failure estimation problem for high-sigma failure rates. In addition, BOIS is described, which was extended in Section III to enable IS without additional explicit circuit simulations. Section IV

978-3-9819263-4-7/DATE20/©2020 EDAA

covers the evaluation results and Section V concludes the paper.

II. Preliminaries

In the following, we briefly review the theory of Importance Sampling to formally introduce the problem and develop necessary notation.

A. Failure Rate Estimation using Monte Carlo

The failure rate of a given circuit, impacted by a process variation model $p(\vec{x})$ can be obtained by the following integral:

$$P_f = \int_{\mathbb{X}} \mathbb{1}(f(\vec{x}) > f_{crit})\, p(\vec{x})\, d\vec{x},$$

where $\mathbb{X}$ denotes the parameter space, and

$$\mathbb{1}(f(\vec{x}) > f_{crit}) = \begin{cases} 1, & \text{if } f(\vec{x}) > f_{crit} \\ 0, & \text{if } f(\vec{x}) \leq f_{crit} \end{cases}$$

denotes the indicator function. Furthermore, $f(\vec{x})$ is a simulation output (e.g. delay, power consumption, gain) and f_{crit} defines the critical threshold above which circuit failures occur. Moreover, $\vec{x}$ is the vector of random variables which impact the simulator output and are induced by process variations (e.g. variations in transistor threshold voltages, capacitances or resistances etc. [9]). Without the loss of generality, we assume that the individual entries x_i i.e. design parameters, are independent identically distributed. Note that correlation in input parameters could be eliminated through Principal Component Analysis (PCA) or Independent Component Analysis (ICA) to achieve a set of uncorrelated random variables. We can therefore model the distribution of the design parameters as identically independent and normal distributed, i.e. $p(\vec{x}) = \mathcal{N}(x; 0, I)$.

As we cannot obtain a closed-form expression for P_f due to $f(\vec{x})$ only being accessible via simulations, P_f is usually estimated using Monte Carlo methods (MC):

$$\begin{aligned} P_f &= \int_{\mathbb{X}} \mathbb{1}(f(\vec{x}) > f_{crit})\, p(\vec{x})\, d\vec{x}, \\ &= \mathbb{E}_{\vec{x}\,\sim\, p(\vec{x})}\left[\, \mathbb{1}(f(\vec{x}) > f_{crit})\, \right] \\ &\simeq \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(f(\vec{x}_i) > f_{crit}), \end{aligned}$$

where P_f is calculated as an expected value of the indicator function and N is the number of samples drawn from a design parameter distribution $p(\vec{x})$.

Since P_f is expected to be extremely small for high-sigma problems, a lot of samples are required for MC methods to converge, which might be computationally infeasible in practice.

B. Importance Sampling

One technique to estimate the probability of rare events is Importance Sampling (IS). In IS, instead of using the distribution $p(\vec{x})$ for sampling, one samples from a different distribution $q(\vec{x}; \vec{\mu}_q)$ instead (see Fig. 3). The distribution $q(\vec{x}; \vec{\mu}_q)$ usually represents a shifted-variant (through $\vec{\mu}_q$) of $p(\vec{x})$ as e.g. in [6], [7], that is closer to the failure region and therefore samples more points $\vec{x}$ surpassing the critical simulator output value i.e. $f(\vec{x}) > f_{crit}$.

Since the samples from $q(\vec{x}; \vec{\mu}_q)$ are drawn under a different distribution, one has to correct for the probability of observing a sample $\vec{x}$ under $p(\vec{x})$. This is done via a correction term $p(\vec{x})/q(\vec{x}; \vec{\mu}_q)$, which is commonly referred to as weight-factor $w(\vec{x})$. The failure rate of the IS estimate can be expressed by

$$\begin{aligned} P_f &= \int_{\mathbb{X}} \mathbb{1}(f(\vec{x}) > f_{crit})\, \frac{p(\vec{x})}{q(\vec{x}; \vec{\mu}_q)}\, q(\vec{x}; \vec{\mu}_q)\, d\vec{x} \\ &= \int_{\mathbb{X}} \mathbb{1}(f(\vec{x}) > f_{crit})\, w(\vec{x})\, q(\vec{x}; \vec{\mu}_q)\, d\vec{x} \\ &= \mathbb{E}_{\vec{x}\,\sim\, q(\vec{x};\vec{\mu}_q)}\left[\, \mathbb{1}(f(\vec{x}) > f_{crit})\, w(\vec{x})\, \right] \\ &\simeq \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(f(\vec{x}_i) > f_{crit})\, w(\vec{x}_i). \end{aligned} \tag{1}$$

Using IS for estimating rare failures is thus a two step process: First, the shift vector $\vec{\mu}_q$ has to be extracted and second, IS has to be carried out on the shifted distribution $q(\vec{x}; \vec{\mu}_q)$. Even though the application of IS is straightforward, the quality (accuracy) of the estimated failure rate strongly depends on the first step, where an appropriate $q(\vec{x}; \vec{\mu}_q)$ needs to be found.

C. Bayesian Optimization

Bayesian Optimization (BO) [10] is a framework for derivative free optimization of expensive to evaluate black-box functions. In the context of Design Automation, it has for example been applied in analog circuit synthesis [11], [12], yield optimization [9] and failure rate estimation [7]. The optimization problem solved by BO can be expressed through:

$$\min_{\vec{x}\in\mathbb{X}} \quad L(f(\vec{x})),$$

where $\mathbb{X}$ is a compact domain. The function $f(\vec{x})$ is thereby some expensive to evaluate black-box function, e.g. a circuit simulation, while $L(z)$ is a so-called loss function judging the favorability of the output of $z = f(\vec{x})$ for a given design parameter configuration $\vec{x}$. In the case of analog circuit synthesis [12], $f(\vec{x})$ can be a certain simulation output, e.g. GAIN of an Operational Amplifier, and $L(z) = -z$ can be used to achieve maximization. In failure rate estimation via BOIS [7], where BO is used to find probable failure points, $L(z) = (z - f_{crit})^2$ is the objective (delay margin), i.e. the quadratic difference between the simulation output $f(\vec{x})$ (e.g. delay) and the critical target f_{crit}.

To solve optimization problems, BO relies on a so called stochastic surrogate model $y_L(\vec{x})$ to iteratively learn about the function $L(f(\vec{x}))$ to be minimized. The points on which the surrogate model is built are thereby sequentially chosen by maximizing a so called acquisition function [13].

a) Gaussian Processes as surrogate models for BO: One frequently used surrogate model $y_L(\vec{x})$ for BO are Gaussian Processes (GP). A GP describes a distribution over functions and is defined as a collection of random variables, where every finite subset is jointly Gaussian distributed [14, Ch. 2]. In the presence of (training) data $D_L = \left\{ \left(\vec{x}_j,\ L(f(\vec{x}_j))\right) \right\}_j^M$, which in the context of BO comes from evaluations of the simulator output, the distribution of $y_L(\vec{x})$ can be updated via

Bayes rule to obtain the posterior distribution $p(y_L(\vec{x}) \mid D_L)$ characterized by its posterior mean $(m_L(\vec{x}))$ and variance $(s_L^2(\vec{x}))$ function [14, Ch. 2].

b) Acquisitions functions for BO: The acquisition function $\alpha\big(\vec{x};\, y_L(\vec{x}), D_L\big)$ is the second ingredient in the optimization procedure of BO.

It is used to select the next point for evaluation given the surrogate model $y_L(\vec{x})$ and previous observations D_L and thereby guides the search for the optimum.

Common choices for acquisition functions are Probability of Improvement (PI), Expected Improvement (EI) and Upper Confidence Bound (UCB) (details see [13]). EI was used in yield optimization [9], while PI was used for failure rate estimation in BOIS [7]. Also, a more specialized acquisition function, namely constraint weighted EI [15] (see also [16]) was used in circuit synthesis to handle performance constraints [11], [12].

D. Bayesian Optimized Importance Sampling (BOIS)

BOIS [7] is a technique for importance sampling based failure rate estimation (of very rare events). It utilizes BO, to identify the failure region via minimizing the objective function (e.g. delay margin), i.e. $L(f(\vec{x})) = (f(\vec{x}) - f_{crit})^2$. Then, an optimization problem is formulated on the resulting surrogate model to extract a shift vector $\vec{\mu}_q$, i.e. mean point, of the IS sampling distribution and IS is performed. The entire procedure of BOIS can be summarized in the following three steps:

1) Create a surrogate model $y_L(\vec{x})$ of the objective function $L(f(\vec{x})) = (f(\vec{x}) - f_{crit})^2$ by minimizing $L(f(\vec{x}))$ with BO.
2) Solve a constrained optimization problem to find the most probable point reaching a specific simulation output $f(\vec{x})$ according to the surrogate model $y_L(\vec{x})$.
3) Perform IS (with circuit simulations) using the optimal point of the optimization problem as shift vector $\vec{\mu}_q$ of of $q(\vec{x}; \vec{\mu}_q)$ and estimate the failure rate.

By combining BO with a local search procedure on the surrogate model, BOIS managed to extract better shift vectors than GIS [6] in direct comparison [7].

III. Extended Bayesian Optimization based Importance Sampling (XBOIS)

Even though BOIS [7] could advance on the state of the art by improving the shift vector extraction, the IS procedure still renders the failure rate estimation costly. In fact, the IS procedure accounts for the majority of the total circuit simulations (and runtime).

To address this major shortcoming, we adapt the idea of [8], where a small set of (uniformly sampled) circuit simulations is used to develop a surrogate model of the simulation output. Through this, expensive circuit simulations can be replaced with cheap surrogate model evaluations in IS.

We propose XBOIS (extended BOIS) as a way to equip BOIS with this capability. However, in contrast to uniformly sampled points, we utilize BO to efficiently guide the development of the surrogate model. A flowchart of the XBOIS methodology can be seen in Fig. 1.

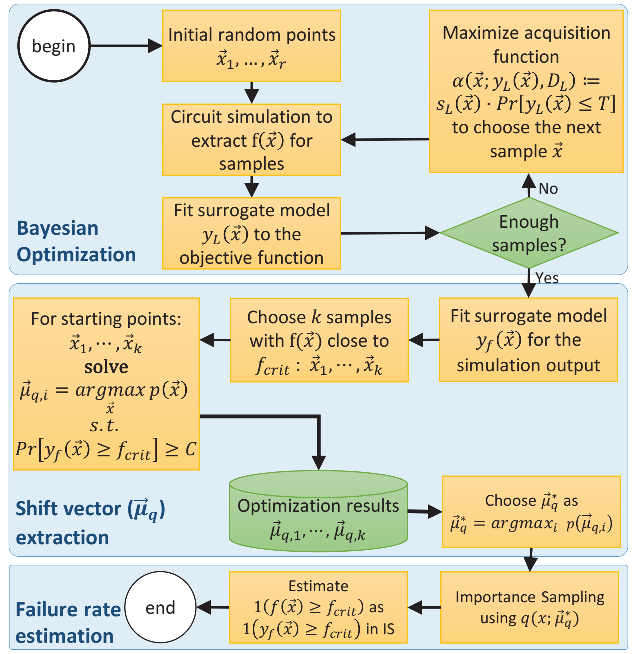

Fig. 1: Overall flow of the XBOIS methodology.

A. A new Acquisition Function for BOIS

As described before, we want to utilize a surrogate model $y_f(\vec{x})$ to estimate the simulation output $f(\vec{x})$ (e.g. delay) for IS. To this end, we require a good approximation of the simulator output $f(\vec{x})$ close to the pass-/fail-border, i.e. the region where we want to evaluate IS samples.

The PI [17] acquisition function used in BOIS is not sufficient for this, as it only tries to minimize the objective function and possibly only samples a small region to refine its found optimum (e.g. Fig. 2). This behaviour is inherently unsuitable for obtaining an approximation of $f(\vec{x})$. Furthermore, PI is known to get stuck in local optima [13]. We therefore propose a more suitable acquisition function that exhibits a guided exploration of the critical simulator output region (i.e. regions with high probability of failure).

The acquisition function has to balance two aspects, namely finding regions with low values of $L(f(\vec{x}))$, as well as exploring the respective region in order to achieve a good approximation of $f(\vec{x})$.

The first aspect can be formalized by maximizing the probability of the next sample $\vec{x}$ having a lower value $L(f(\vec{x}))$ than a certain user define threshold T. This can be estimated by $Pr\big[y_L(\vec{x}) \leq T\big]$ using the surrogate model $y_L(\vec{x})$.

The second aspect, namely achieving a good approximation of the objective function surface, can be formulated as minimizing the uncertainty of $y_L(\vec{x})$. One measure of the uncertainty of $y_L(\vec{x})$ is the Differential Entropy (see [14, App. A]), which is proportional to the posterior standard deviation $s_L(\vec{x})$ of $y_L(\vec{x})$ at $\vec{x}$. We can combine both aspects in the following acquisition function:

$$\alpha\big(\vec{x}; y_L(\vec{x}), D_L\big) = s_L(\vec{x}) \cdot Pr\big[y_L(\vec{x}) \leq T\big]$$

The multiplication of both terms leads to the acquisition function exhibiting high values in regions where both terms are high. This way, previously unexplored regions close to the

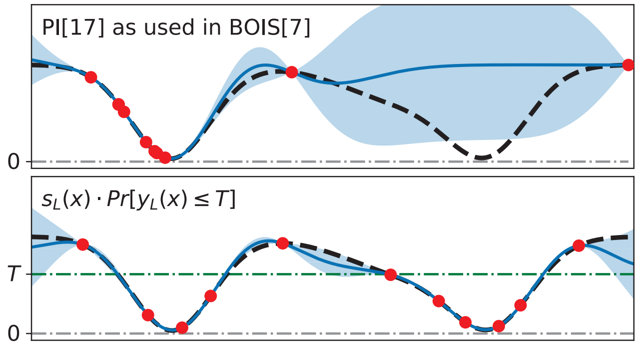

Fig. 2: An example of the behaviour of PI [17] versus the proposed acquisition function. The mean function $m_L(\vec{x})$ and the 1-standard deviation region $s_L(\vec{x})$ of $y_L(\vec{x})$ are displayed in blue and light blue respectively. The dotted black line indicates the true unknown $L(f(\vec{x}))$. Each plot shows 11 points, i.e. 1 random and 10 chosen by the respective acquisition function.

pass-/fail-border are favoured, increasing our knowledge about design configurations $\vec{x}$ with simulation outputs $f(\vec{x})$ close to f_{crit}. See Fig. 2 for a comparison of PI, as used in BOIS, vs. the proposed acquisition function. Note that PI tries to refine the minimum of $L(f(\vec{x}))$ while potentially ignoring regions of similar low values. In contrast, the proposed acquisition function leads to a better approximation of the objective in the region below T. The refinement of the minimum is thereby less enforced. Alternatively, the construction of the acquisition function can also be understood as maximizing $s_L(\vec{x})$ under the condition that $L(f(\vec{x})) \leq T$ where the constraints are handled as proposed in [15], [16].

B. Shift Vector Extraction and Importance Sampling

To extract the shift vector, i.e. mean vector $\vec{\mu}_q$ of the sampling distribution $q(\vec{x}; \vec{\mu}_q)$, we first build a surrogate model $y_f(\vec{x})$ of the simulation output $f(\vec{x})$ based on the M collected data points $D_f = \left\{\left(\vec{x}_j,\ f(\vec{x}_j)\right)\right\}_j^M$. Then, similar to BOIS [7], the shift vector $\vec{\mu}_q$ is extracted:

$$\vec{\mu}_q = \underset{\vec{x}}{argmax} \quad p(\vec{x}) \quad \text{s.t.} \quad Pr\left[y_f(\vec{x}) \geq f_{crit}\right] \geq C,$$

where C is a user defined confidence level. Following the shift vector extraction, we can perform IS using $q(\vec{x}; \vec{\mu}_q)$:

$$\begin{aligned} P_f &\simeq \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(f(\vec{x}_i) > f_{crit})\, \frac{p(\vec{x}_i)}{q(\vec{x}_i; \vec{\mu}_q)} \\ &\simeq \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(y_f(\vec{x}_i) > f_{crit})\, \frac{p(\vec{x}_i)}{q(\vec{x}_i; \vec{\mu}_q)}, \end{aligned}$$

where the samples x_i are drawn from $q(\vec{x}_i; \vec{\mu}_q)$. Note that we evaluate $y_f(\vec{x}_i)$ in the IS step, i.e. the surrogate model of the simulation output $f(\vec{x})$, instead of performing costly circuit simulations to obtain $f(\vec{x}_i)$.

An illustration of the full procedure on a two dimensional 6T-SRAM cell array subproblem can be seen in Figure 3. Through the use of the specialized acquisition function, BO samples points close to the critical delay f_{crit} and therefore achieves a good approximation of $f(\vec{x})$ in this region.

TABLE I: Three scenarios: Operation modes of 6T-SRAM cell

	VDD	f_{crit}	#Transistors	Failure rate	Golden
Scenario 1	0.5V	10ns	12	Low Sigma	MC10000
Scenario 2	0.5V	16ns	2	High Sigma	GridSearch
Scenario 3	1.0V	350ps	12	High Sigma	-

IV. Experiments

In this section, we apply the proposed methodology to a failure rate estimation problem of a 6T-SRAM cell array, where the simulator output is the read latency time (RLT) (or delay) of a SRAM cell array read operation (see Fig. 4). Moreover, failures are defined as RLT violations, where the delay exceeds a pre-defined limit (f_{crit}). The proposed method is compared versus two state-of-the-art approaches: BOIS [7] and GIS [6], which have been shown to be superior to other related work. As evaluation metrics, we report the accuracy and the total number of circuit simulations.

A. Setup

The aforementioned 6T-SRAM column with 256 word lines (cells) is implemented in Cadence Virtuoso environment based on the 28nm FDSOI library. The circuit schematic is depicted in Fig. 4. The RLT, which is an important circuit characteristic for SRAMs, was extracted using the accurate Spectre circuit simulator. Process variations are modeled as variations in the threshold voltage of the transistors in the circuit. For this analysis, the worst case was considered, where 255 cells store logical '0', while logical '1' is read from a target cell [6]. For this setup, the 6T-SRAM column can be simplified by combining the 255 cells into a virtual SRAM-cell with 255 times wider channel width to account for the read disturb. This virtual SRAM cell, the target cell as well as the sense amplifier, which performs the read operation, are illustrated in Fig. 4. This leads to 12 parameters with variation from the transistors of this circuit, which are assumed to be normal distributed and mutually independent [6].

B. Evaluation

In order to prove the effectiveness of the proposed approach, we evaluated the XBOIS method on different scenarios. The configuration parameters of each scenario are given in Table I.

For the following analysis, we set the user-defined threshold to $T = 2$. The supply voltage and critical delay (i.e. f_{crit}) was varied to obtain low-sigma (rare events) and high sigma (very rare events) problems. In addition, the number of transistors under variation was varied to get low-dimensional (< 10 transistors) or high-dimensional (≥ 10 transistors) scenarios. For the golden result extraction, we used exhaustive MC simulations (10000 simulations) for Scenario 1, and grid-search for Scenario 2 (see Table I). For Scenario 3, the golden result cannot be obtained in practice (due to high-sigma and high-dimensions), however comparison between methods can still be accomplished, due to the fact that more accurate IS methods are always associated with higher failure rate estimates (which is automatically expressed through (1), as the weights $w(x)$ become small for non-optimal shift vectors). For

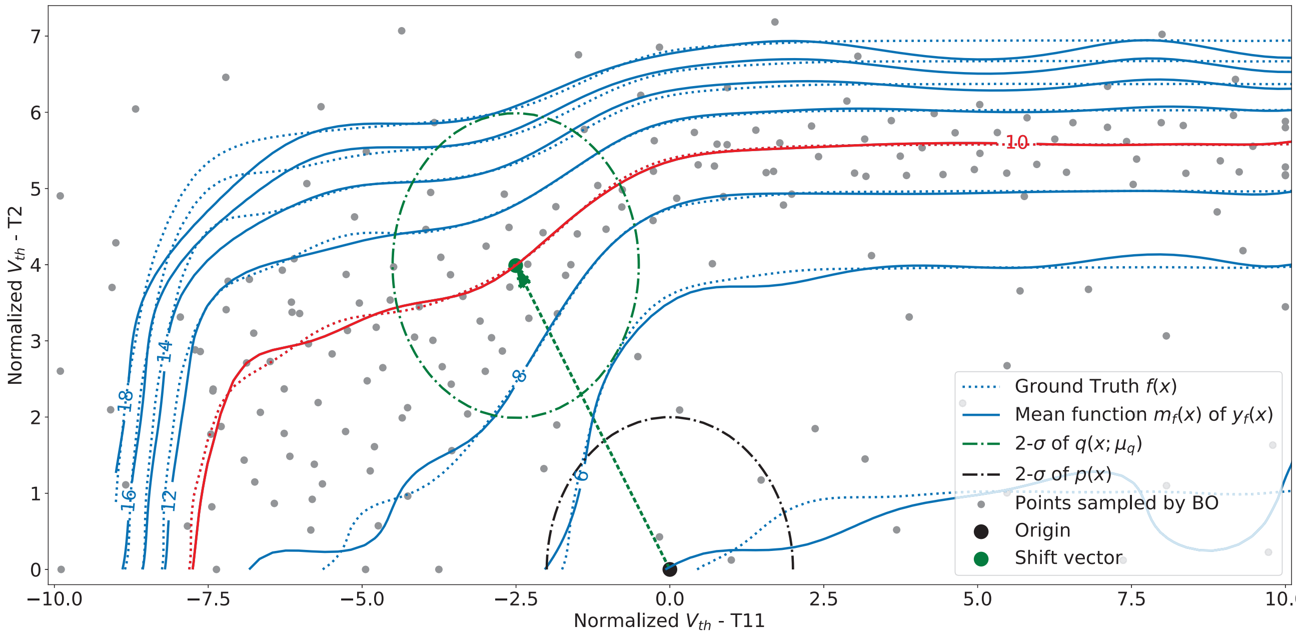

Fig. 3: An illustration of the XBOIS procedure on a two dimensional 6T-SRAM cell array subproblem. In this example the critical delay was set to 10ns (red)(pass-/fail-border). The design parameter T in the acquisition function was set to $T = 2$. The black and green circle (ellipse due to axis scaling) denote the 2-sigma level-set of the distributions $p(\vec{x})$ and $q(\vec{x}; \vec{\mu}_q)$ respectively. To strongly emphasize the sampling behaviour of the acquisition function around the critical delay, we used M=200 samples (grey points) to build the surrogate model $y_f(\vec{x})$ of the simulator output (delay), note that fewer are sufficient.

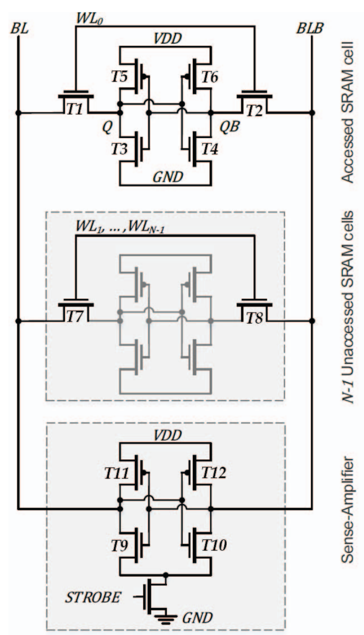

Fig. 4: 6T-SRAM and sense-amplifier in a memory column. In total 12 transistors (T1-T12) under variation for high-dimensional problem (12D) or 2 transistors (T2, T11) for low-dimensional problem (2D).

this reason, we used the accurate XBOIS result as the basis of comparison.

We compared the proposed approach and state-of-the-art techniques by assessment of the final failure rate (P_f) estimate obtained after running a pre-determined number of circuit simulations.

TABLE II: Comparison of XBOIS, BOIS and GIS (accuracy, runtime)

	Runs	XBOIS	BOIS [7]	GIS [6]
Scenario 1 (Low-Sigma, High-Dimensional)	Shift Vector	600	200	200
	IS	0	800	800
	Total	600	1000	1000
	Relative Error	1.1%	1.5%	6.6%
	Runs	XBOIS	BOIS [7]	GIS [6]
Scenario 2 (High-Sigma, Low-Dimensional)	Shift Vector	600	400	400
	IS	0	3600	3600
	Total	600	4000	4000
	Relative Error	22.7%	58.4%	99.7%
	Runs	XBOIS	BOIS [7]	GIS [6]
Scenario 3 (High-Sigma, High-Dimensional)	Shift Vector	600	200	200
	IS	0	9800	9800
	Total	600	10000	10000
	Underestimation	1x	589x	$\sim 3.5{\cdot}10^5$x

C. Summary

The simulation results are summarized in Table II. For each scenario, the breakdown of the total number of evaluations into shift vector extraction and IS is given. According to [7], the number of IS simulations was set to 3600 or 9800, respectively. In Table II also the relative error to the golden result is mentioned. For Scenario 3, the underestimations of the failure rate of [6], [7] compared to XBOIS are shown.

For the first two scenarios, the relative error of our proposed approach against the golden result was reduced significantly (Scenario 1) or substantially (Scenario 2). As can be obtained

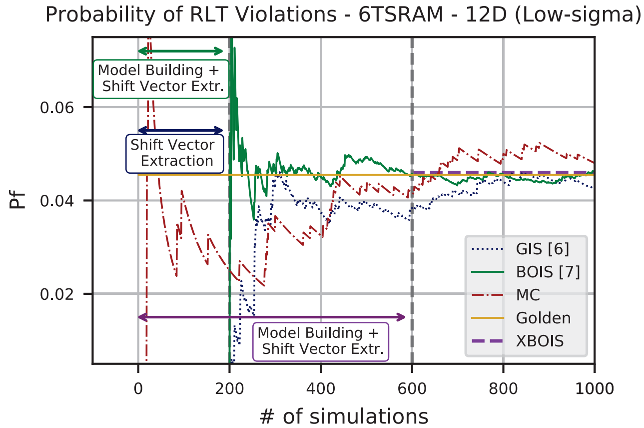

Fig. 5: Scenario 1: Failure rate P_f estimates for XBOIS, BOIS [7], GIS [6], Monte Carlo (MC) and Golden result.

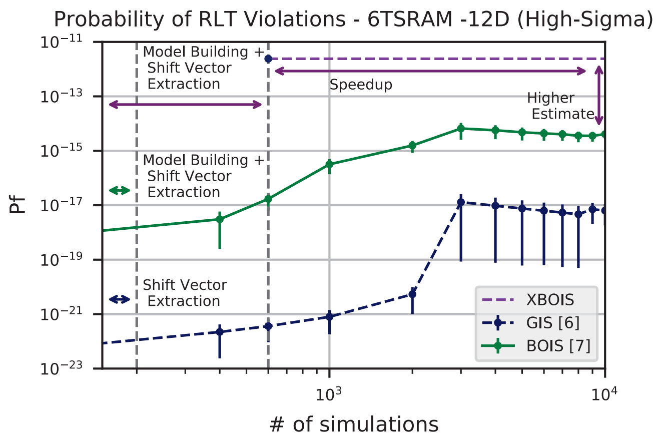

Fig. 6: Scenario 3: Comparison of failure rates for all approaches: GIS [6], BOIS [7] and XBOIS against the number of required circuit simulations. XBOIS reaches substantial speedup (faster convergence) as well as higher failure rate estimate. All plots with error bars. Failures defined as read latency time (RLT) violations.

from Fig. 5, after 600 simulations XBOIS is terminated and returns a failure rate estimate very close to the golden result, while other approaches still require more time to converge.

In Scenario 3, which is closest to a real world application, the failure rate underestimation compared to XBOIS is about $3.5{\cdot}10^5\times$, and $589\times$ for GIS [6] and BOIS [7], respectively. This indicates that a better shift vector was found. Secondly, the required number of circuit simulations are reduced by $16.6\times$ (from 10000 to 600) as XBOIS uses the surrogate model for IS.

The comparison of failure rate estimations evolving over circuit simulations is also illustrated in Fig. 6. For XBOIS 600 simulations are sufficient to extract a two orders of magnitude higher failure rate estimate compared to BOIS. The speedup of XBOIS compared to BOIS and GIS is $10000/600 = 16.6$.

V. Conclusion

In this work, we propose a novel Importance Sampling (IS) technique, which is capable of solving the crucial failure rate estimation problem of rare events using very few circuit simulations. The underlying failure rate estimation method is based on Bayesian Optimized Importance Sampling, adapted to explore the problem space more efficiently. This allows for the construction of an accurate surrogate model of the simulator output, which can substitute the circuit simulations required in IS. Through this, the computational effort of the failure rate estimation is reduced substantially, achieving speedups of up to 16x.

Acknowledgements

This work was supported by the Ministry of Science, Research and Arts of the state of Baden-Württemberg in form of the MERAGEM doctoral program.

References

[1] S. R. Nassif, "Design for variability in dsm technologies [deep submicron technologies]," in *Proceedings IEEE 2000 First International Symposium on Quality Electronic Design (Cat. No. PR00525)*. IEEE, 2000, pp. 451–454.

[2] R. Heald and P. Wang, "Variability in sub-100nm sram designs," in *Proceedings of the 2004 IEEE/ACM International conference on Computer-aided design*. IEEE Computer Society, 2004, pp. 347–352.

[3] S. Sun *et al.*, "Fast statistical analysis of rare circuit failure events via scaled-sigma sampling for high-dimensional variation space," in *Proceedings of the International Conference on Computer-Aided Design*. ACM, 2013, pp. 478–485.

[4] 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 *2006 43rd ACM/IEEE Design Automation Conference*. IEEE, 2006, pp. 69–72.

[5] L. Dolecek *et al.*, "Breaking the simulation barrier: SRAM evaluation through norm minimization," in *ICCAD*, 2008.

[6] T. Haine *et al.*, "Gradient importance sampling: An efficient statistical extraction methodology of high-sigma sram dynamic characteristics," in *DATE*, 2018.

[7] D. D. Weller *et al.*, "Bayesian optimized importance sampling for high sigma failure rate estimation," in *2019 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. IEEE, 2019, pp. 1667–1672.

[8] M. Malik *et al.*, "Sparse regression driven mixture importance sampling for memory design," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 26, no. 1, pp. 63–72, 2018.

[9] M. Wang *et al.*, "Efficient Yield Optimization for Analog and SRAM Circuits via Gaussian Process Regression and Adaptive Yield Estimation," *TCAD*, 2017.

[10] J. Močkus, "On bayesian methods for seeking the extremum," in *Optimization Techniques IFIP Technical Conference Novosibirsk, July 1–7, 1974*, G. I. Marchuk, Ed. Berlin, Heidelberg: Springer, 1975, pp. 400–404.

[11] W. Lyu *et al.*, "An efficient bayesian optimization approach for automated optimization of analog circuits," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 65, no. 6, pp. 1954–1967, June 2018.

[12] S. Zhang *et al.*, "Bayesian optimization approach for analog circuit synthesis using neural network," in *2019 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. IEEE, 2019, pp. 1463–1468.

[13] B. Shahriari *et al.*, "Taking the human out of the loop: A review of bayesian optimization," *Proceedings of the IEEE*, vol. 104, no. 1, pp. 148–175, Jan 2016.

[14] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, 2nd ed. MIT Press, 2006.

[15] M. A. Gelbart, J. Snoek, and R. P. Adams, "Bayesian optimization with unknown constraints," *arXiv preprint arXiv:1403.5607*, 2014.

[16] J. R. Gardner *et al.*, "Bayesian optimization with inequality constraints." in *ICML*, 2014, pp. 937–945.

[17] H. J. Kushner, "A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise," *Journal of Basic Engineering*, vol. 86, no. 1, pp. 97–106, 1964.