NPC-NIS: Navigating Semiconductor Process Corners with Neural Importance Sampling

Hong Chul Nam HONGCHUL.NAM@ALSEMY.COM
Research & Development Center
Alsemy Inc.
Seoul, Korea

Chanwoo Park CHANWOO.PARK@ALSEMY.COM
Research & Development Center
Alsemy Inc.
Seoul, Korea

Abstract

Traditional corner case analysis in semiconductor circuit design typically involves the use of predetermined semiconductor process parameters, including Fast, Typical, and Slow corners for PMOS and NMOS devices, frequently yielding overly conservative designs due to the utilization of fixed, and potentially non-representative, process parameter values for circuit simulations. Identifying the worst cases of circuit FoMs within typical semiconductor process variation ranges presents a considerable challenge, especially given the complexities associated with accurately sampling rare semiconductor events. In response, we introduce NPC-NIS, a model specifically developed for estimating rare cases in semiconductor circuit analysis, leveraging a learnable importance sampling strategy. We model the distribution of process parameters that exhibit the worst FoMs within a realistic range. This adaptable framework dynamically identifies and addresses rare semiconductor cases within typical process variation ranges, enhancing our circuit design optimization capabilities under realistic conditions. Our empirical results validate the effectiveness of the Neural Importance Sampling (NIS) approach in identifying and mitigating rare semiconductor scenarios, thereby contributing to the development of more robust and reliable semiconductor circuit designs and connecting traditional semiconductor corner case analysis with real-world semiconductor applications.

Keywords: Corner Case, Reinforcement Learning, Importance Sampling

1. Introduction

Traditional semiconductor design, focused on optimizing performance, power, and area, encounters notable constraints when utilizing corner case analysis (Weste and Harris, 2015). The current methods, which use set semiconductor process parameters for simulation, often impose strict and potentially unrealistic conditions for design scenarios. These fixed corners ensure designs work under extreme scenarios but don't typically represent most real-world process variations, leading to overly safe and pessimistic designs. Moreover, as more process variables are included, the complexity of simulations grows, demanding prohibitive computational resources and time. This creates a clear need for a method that acknowledges real and varied process changes and allows for efficient simulation, particularly in high-

dimensionality situations, avoiding the computational struggles and restrictions of existing strategies.

Park et al. (2023) proposed a method for density estimation in yield and high-sigma analysis with fewer simulations to manage high-dimensional data without the necessity for complex training of relationships between input and output pairs of SPICE simulations. Studies such as those by Amrouch et al. (2020) on Negative Capacitance FinFET technology, Pond et al. (2017) on Monte Carlo analysis of photonic integrated circuits, and Rakka et al. (2020) on hybrid Importance Sampling methodologies have provided valuable insights into the performance of technologies under process variations. While these studies contribute significantly to the field, they focus on different aspects of variability and do not specifically target the challenges of corner case analysis and the development of a learnable sampling method.

Importance Sampling (IS) has been widely utilized in simulations to mitigate integration bias, particularly in enhancing the sampling of extreme distributions. This research pioneers the incorporation of Neural Importance Sampling (NIS) into process variation modeling, aiming to increase the sampling of corner processes within a typical range, and demonstrates that this method is equivalent to policy gradient. Moreover, it introduces a unique approach that, instead of defining process corners, which can result in pessimistic designs, adapts to individual circuit contexts. Specifically, in scenarios where we lack knowledge of the true worst combinations of process variables for circuit Figures of Merit (FoMs), the traditional brute-force simulation becomes increasingly impractical as the dimensionality of process variables escalates into the thousands, thus encountering the Curse of Dimensionality (CoD).

2. NPC-NIS: Navigating Process Corners with Neural Importance Sampling

Unlike traditional corner models which define Fast and Slow corners for PMOS and NMOS devices and perform estimation over those given parameters, NPC-NIS does not explicitly define the boundary for worst-case process parameters. There are multiple reasons for this. First, traditional corner values do not necessarily guarantee the worst performance of circuits as extreme cases of certain parameters do not necessarily imply the worst case. Second, traditional corner models perform very pessimistic estimations where the probability of fabrication over those points is extremely unlikely. Third, corner definitions require also domain expert knowledge, making it difficult for new engineers to learn necessary corner values.

FoM is calculated as a weighted sum of normalized performance metrics as shown in Equation 1:

$$\mathrm{f(x)} = \sum_{i=1}^{n} w_i \frac{SPICE(x)_i - y_i^{min}}{y_i^{max} - y_i^{min}} \tag{1}$$

where $SPICE$ is a SPICE circuit simulator . Above FoM is used as a performance metric where y_{min} and y_{max} are obtained from 10,000 randomly sampled circuit designs and used for metric normalization. w_i is a weighting factor that determines whether a metric has to be maximized or minimized for the better FoM. Hereby we assume that all metrics need to be minimized (e.g. power consumption, delay).

2.1 Adaptive Importance Sampling

We define our objective function as

$$\mu = \mathbb{E}_{x\sim p}[\mathbb{1}_{\{f(x)>\gamma\}}] = \mathbb{P}[f(x) > \gamma] \tag{2}$$

where γ is a failure threshold, i.e. any fabrication with metrics above this threshold level is considered a failure, and $p \sim \mathcal{N}(\mu, \Sigma)$ is a probability distribution of each process variable.

Monte-Carlo (MC) simulation performs the modeling of this probability by estimating an empirical estimation such that

$$\hat{\mu} = \frac{1}{N}\Sigma_{i=1}^{N}\mathbb{1}_{\{f(x_i)>\gamma\}} \tag{3}$$

We define our weight function as

$$k(x) = \frac{p(x)}{q_\theta(x)} = \frac{\mathcal{N}(\mu, \Sigma)}{q_\theta(x)} \tag{4}$$

Here the nominator of the weight function is the pdf of a Gaussian distribution, which we define as a normal distribution over each process variable. This term $k(x)$ is used to compensate for the sampling due to the shift of sampling distribution.

2.2 Neural Importance Sampling

An optimal proposal distribution $q^*(x)$ can be represented as

$$q^*(x) = \frac{\mathbb{1}_{\{f(x)>\gamma\}}p(x)}{\mu} \tag{5}$$

However, this is in practice impossible due to the requirement of the knowledge of μ. As a result, we approximate $q^*(x)$ by a tractable approximator $q_\theta(x)$ parameterized by θ. Multiple models are possible for having the proposal for NIS, we use normalizing flow (Papamakarios et al., 2021) as our function class for estimating it.

Based on the analysis, we perform the cross-entropy method (CEM) by minimizing the KL divergence between the optimal and parametrized distributions to get the optimal parameter $\theta = \arg\min_\theta D_{KL}(q^*||q_\theta)$.

We define our policy objective function

$$J(\theta) = \mathbb{E}_{x\sim q_\theta}[k(x)\mathbb{1}_{\{f(x)>\gamma\}}\log q_\theta(x)] \tag{6}$$

3. Experiment

We perform experiments over 5-ring oscillators (RO) with SPICE simulator. We compare NPC-NIS with MC simulations and Gaussian importance sampling (GIS). We adaptively learn the mean and variance of Gaussian samples by performing gradient descent with the same loss function as NIS and assume no correlation between each individual process parameter to reduce the number of parameters.

We ensure that NPC-NIS and GIS are well-trained by tuning hyperparameters over grids and picking the best-performing parameters.

Algorithm 1 NPC-NIS

Require:

θ_0: initial parameter values
N_{total}: total number of iterations
B: batch size for sampling
γ: threshold level
ρ: quantile level

0: **for** i $< N_{\text{total}}$ **do**
0: Sample $\{x_i\}_{i=1}^{B}$ from q_{θ_k}
0: Estimate $f(x_i)$ for each x_i
0: Sort $\{f(x_i)\}_{i=1}^{B}$ such that $f(x_i) \leq f(x_{i+1})$
0: Compute $\gamma_k = \max(\gamma, f(x_\rho))$
0: Update θ_k by $\theta_{k+1} = \theta_k + \nabla_\theta J(\theta)$
0: $i = i + 1$
0: **end for**=0

3.1 Sampling Range

We define our ranges for process parameters as shown in Table 1. To simulate a more realistic fabrication scenario, we define min/max points as our 5-sigma points with typical values as means.

Table 1: **Process variable ranges for NMOS and PMOS.**

Process Variable	Symbol	Unit	NMOS Range	PMOS Range
Oxide Thickness	t_{ox}	nm	1.53–2.07	1.55–2.09
Threshold Voltage	V_{th0}	V	0.498–0.747	-0.704–-0.470
Velocity Saturation	v_{sat}	m/s	104,000–156,000	72,000–108,000
Mobility	μ_0	m^2/Vs	0.0343–0.0637	0.0147–0.0273
Subthreshold Swing	N_{factor}	–	1.44–1.76	1.62–1.98

4. Results

We compare NPC-NIS with Monte Carlo simulation (MC) and Gaussian Importance Sampling (GIS) over the proportion of points sampled at different sigma intervals after running 10,000 simulations with each batch size of 10.

Table 2 shows the number of points between each sigma quantile, where each quantile value is determined by performing 5×10^7 random simulations. The performance shows that NPC-NIS is able to sample more points at high-σ points ($\geq 3\sigma$) in comparison to every other model. In particular, it was able to sample points at 6σ regions, which is a particularly challenging task when the probability of sampling a point at 6σ equals $1.973 \times 10^{-7}\%$.

Table 3 shows the cumulative number of points sampled at each σ region, and the result shows that NPC-IS samples at extreme corners more efficiently at every cumulative σ interval.

Table 2: **Counts of points lying on each inidivudal sigma region where each sigma region is defined by quantiles from 5×10^7 simulations by running 10,000 simulations**

Type	$0-1\sigma$	$1-2\sigma$	$2-3\sigma$	$3-4\sigma$	$4-5\sigma$	$5-6\sigma$	$6-7\sigma$
MC	**2919**	**5479**	844	52	2	0	0
GIS	658	2392	**1901**	1608	1229	290	0
NPC-NIS	638	2336	1868	**1662**	**1272**	**362**	**1**

Table 3: **Cumulative counts of points belonging to each σ region**

Type	1σ	2σ	3σ	4σ	5σ	6σ
MC	6377	898	54	2	0	0
GIS	7420	5028	3127	1519	290	0
NPC-IS	**7501**	**5165**	**3297**	**1635**	**363**	**1**

5. Discussion

If one is familiar with policy gradient (Williams, 1992), one can see that the objective is very similar to policy gradient where

$$J(\theta) = \mathbb{E}_{x\sim q_\theta}[k(x)\mathbb{1}_{\{f(x)>\gamma\}}] \tag{7}$$

Policy gradient performs gradient ascent by

$$\nabla_\theta J(\theta) = \mathbb{E}_{x\sim q_\theta}[k(x)\mathbb{1}_{\{f(x)>\gamma\}}\nabla \log q_\theta] \tag{8}$$

and we perform

$$\theta_{t+1} = \theta_t + \nabla J(\theta_t) \tag{9}$$

This form is equivalent to neural importance sampling where for us q_θ is defined as a normalizing flow generative network. This connection helps us to bring in all useful improvements based on policy gradient into neural importance sampling to stabilize the training of the proposal distribution for which rare events are more sampled. Furthermore, we would like to introduce trust region (TR) into future work to make training of our proposal distribution more stable.

6. Conclusion

We introduced NPC-NIS, a model that employs Neural Importance Sampling (NIS) to adeptly handle the complexities of process variation modeling, particularly by sampling process parameters that induce circuits to reach their extreme Figures of Merit (FoMs) without adhering to predefined worst-case corners. Instead of conventional strategies that often lead to conservative designs due to their reliance on fixed process corners, NPC-NIS adopts a more flexible and individualized approach, enabling efficient sampling across wide 1-6 sigma regions and demanding minimal simulation iterations. Consequently, NPC-NIS contributes a notable advancement in connecting traditional semiconductor corner case analysis with tangible, real-world semiconductor applications.

References

Hussam Amrouch, Girish Pahwa, Amol D Gaidhane, Chetan K Dabhi, Florian Klemme, Om Prakash, and Yogesh Singh Chauhan. Impact of variability on processor performance in negative capacitance finfet technology. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 67(9):3127–3137, 2020.

George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *The Journal of Machine Learning Research*, 22(1):2617–2680, 2021.

Chanwoo Park, Hong Chul Nam, Jihun Park, and Jongwook Jeon. Flowsim: An invertible generative network for efficient statistical analysis under process variations. In *2023 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, pages 157–160, 2023.

James Pond, Jackson Klein, Jonas Flückiger, Xu Wang, Zeqin Lu, Jaspreet Jhoja, and Lukas Chrostowski. Predicting the yield of photonic integrated circuits using statistical compact modeling. In *Integrated Optics: Physics and Simulations III*, volume 10242, pages 146–158. SPIE, 2017.

Mariam Rakka, Rouwaida Kanj, and Ragheb Raad. Hybrid importance splitting importance sampling methodology for fast yield analysis of memory designs. In *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5. IEEE, 2020.

Neil HE Weste and David Harris. *CMOS VLSI design: a circuits and systems perspective*. Pearson Education India, 2015.

Ronald J Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8:229–256, 1992.