FlowSim: An Invertible Generative Network for Efficient Statistical Analysis under Process Variations

Chanwoo Park[1*], Hong Chul Nam[1,2*], Jihun Park[1], Jongwook Jeon[1,3]
[1]Research & Development Center, Alsemy Inc., Seoul, Korea
[2]Department of Information Technology and Electrical Engineering, ETH Zurich, Switzerland
[3]School of Electronic and Electrical Engineering, Sungkyunkwan University, Suwon, Korea

***Abstract*—The analysis of statistical variation in circuits or devices, resulting from process, voltage, and temperature (PVT) variations, is a critical aspect of ensuring high yield and accurate high-sigma analysis in semiconductor fabrication. As the industry progresses toward nanometer technologies, process variation becomes a significant challenge, necessitating the development of effective statistical models. Traditional Monte Carlo simulations, however, struggle to scale with the increasing number of process variables, leading to an exponential growth in the required simulations. In response to this challenge, we introduce FlowSim, a novel approach that employs density estimation to accurately perform yield and high-sigma analysis with a significantly reduced number of simulations. This approach offers a unique solution to the scalability issues faced by conventional Monte Carlo simulations, providing over a 100x decrease in the number of required simulations while maintaining a prediction error below 5% across all statistical metrics of circuit performance.**

***Index Terms*—Statistical variation, Monte Carlo simulation, High-sigma analysis, Artificial neural network**

I. Introduction

In recent years, the manufacturing of semiconductors has become increasingly complex due to the ongoing downscaling, leading to smaller feature sizes and tighter process tolerances. This complexity has given rise to process variations, which have emerged as a significant concern due to their potential to cause deviations in circuit performance and yield. This issue is particularly pronounced in advanced technologies where the number of process variables increases. Traditional Monte Carlo simulations have encountered scalability challenges because of these factors.

To address these challenges, we propose FlowSim, to our knowledge, the first approach that utilizes density estimation to achieve accurate yield and high-sigma analysis with a significantly reduced number of simulations. Although surrogate model-based approaches using simple neural networks [4] exist, FlowSim offers distinct advantages in terms of interpretability and managing high-dimensional data. Specifically, FlowSim avoids the need to train complex relationships between input and output pairs of SPICE simulations, which would require large amounts of data as the number of input process variables increases.

The effectiveness of the FlowSim method is demonstrated through its application to ring oscillators (RO) of various stages. The approach achieves notably low error rates: 0.9% and 0.8% for 3 and 4 sigma values of Figures of Merit (FoMs) respectively. Moreover, it estimates the yield of an RO with error rates of 0.2% and 0.5% across different specifications. These experimental results suggest that this machine learning-based technique can potentially overcome the computational challenges associated with traditional Monte Carlo simulations, thereby enabling more precise and efficient statistical analysis of circuit performance under PVT variations.

*Equal contribution. Email: chanwoo.park, hongchul.nam@alsemy.com

II. Related Work

A. Efficiency Improvement in Simulation

Significant research has been conducted to enhance the efficiency and precision of simulations and statistical analyses in the context of process variability. A study presents a simulation framework for evaluating the impact of location-dependent variability in photonic integrated circuits through Monte Carlo simulations, emphasizing the importance of variability modeling at the circuit level for efficient design [3]. Another research proposes an automated framework for Variability Analysis of CMOS circuits using a simulated annealing algorithm, highlighting the increasing impact of process manufacturing and design mismatches on circuit performance [5]. Other works introduce a yield optimization method based on Bayesian optimization [14], a scaled-sigma sampling method for estimating rare failure rates [13], and a new sampling scheme that combines the benefits of Latin Hypercube and Low Discrepancy sampling methods [9].

B. Variability Analysis

Research has also focused on applying various methods to specific technologies, providing valuable insights into their performance under process variations. One paper investigates the impact of Negative Capacitance FinFET technology on processor performance under process variations, revealing that process variations have a larger impact on the processor's performance when it operates at a lower voltage [1]. Another study presents two approaches that enable Monte Carlo analysis of photonic integrated circuits, including the treatment of spatial correlations [11]. A different paper proposes a hybrid Importance Sampling Importance Splitting methodology for rare fail event estimation of high-dimensional memory designs, demonstrating a 3-5X reduction in runtime compared to traditional Importance Splitting approaches [12]. Lastly, a study proposes a low-cost and accurate yield estimation procedure for compact microwave couplers, demonstrating the effectiveness of variable-fidelity electromagnetic (EM) simulation models and fast surrogates [8].

C. Generative AI in Electronic Design Automation

Generative AI has shown promise in various aspects of electronic design automation (EDA). One study introduces a generative adversarial network (GAN)-guided well generation framework for analog/mixed-signal circuit layout, which mimics the behavior of experienced designers and improves layout compactness and robustness [15]. Another paper proposes a deep learning strategy combining wavelet transform and GANs for analog-circuit fault diagnosis, addressing limitations of traditional methods and limited training samples [6]. A different study evaluates the effectiveness of generative self-supervised learning for combinational gate sizing in VLSI designs, achieving high accuracy and significantly faster design convergence [10]. A paper proposes an end-to-end lithography modeling framework based on a GAN, which significantly speeds up the lithography simulation process compared to conventional methods [16]. Lastly, a fully automated analog routing paradigm that leverages machine learning for routing guidance is introduced, achieving significant improvements over existing techniques and competitive performance to manual layouts [17].

While AI tools have been extensively applied to enhance efficiency across various aspects of the electronic design process, few studies have explored direct modeling of circuit performance through density estimation. In particular, the typical regression models, which directly map process variables to figures of merit, often struggle with efficiency and accuracy when dealing with complex high-dimensional dependencies and non-linear relationships. Additionally, while generative models have been used in different contexts, their application as a method for studying the statistical distributions of circuit behaviors due to process variability has not been thoroughly investigated.

III. Generative Model for Density Estimation

FlowSim is a generative model based on Normalizing Flows (NF) [7], designed to estimate the target density of circuit figures of merit (FoM) such as power and delay, influenced by process variations. One of the main advantages of NFs over other generative models, such as generative adversarial networks (GANs) and variational autoencoders (VAEs), is their ability to provide explicit density estimation. This capability is facilitated by a series of invertible transformations, which map data from a simple prior distribution, such as a Gaussian distribution, to the complex target distribution. The result is an immediate and insightful understanding of the data density, which can be beneficial in guiding design decisions and optimization strategies.

As shown in Figure 1, FlowSim starts by training on a set of SPICE simulation results, with quantities ranging from 100 to 1,000,000. The training process aims to minimize the Kullback-Leibler (KL) divergence between the learned distribution and the target distribution, ensuring the model accurately captures circuit variability as reflected in the FoM distribution and dispersion.

After training, FlowSim is able to rapidly generate a large number of synthetic samples that closely resemble the SPICE simulation results. This generation process only requires the inference time of the neural network, which is significantly less computationally demanding compared to SPICE simulations, thus enabling efficient statistical analyses of the circuit. The detailed methodology of FlowSim is outlined in Algorithm 1.

Algorithm 1 Estimating statistics of a circuit using FlowSim

Require: Circuit (e.g., 51-stage RO), SPICE simulator, $q_\theta(x)$, where x denotes FoMs of the circuit;

1: Simulate the circuit with SPICE and generate k samples, $\{x_i\}_{i=1}^k$;

2: Train $q_\theta(x)$ to approximate true density $p(x)$ by minimizing the KL divergence $L(\theta) = \sum_{i=1}^k \log\left(\frac{q_\theta(x_i)}{p(x_i)}\right)$;

3: Generate n samples ($n \gg k$) with $q_\theta(x)$, $\{\tilde{x}_i\}_{i=1}^n$;

4: Compute estimated statistics (e.g., μ, σ, yield) using synthetic data $\{\tilde{x}_i\}_{i=1}^n$;

IV. Experiments and Results

We conducted experiments on 5, 51, and 101-stage Ring Oscillators (ROs) to demonstrate efficient density estimation

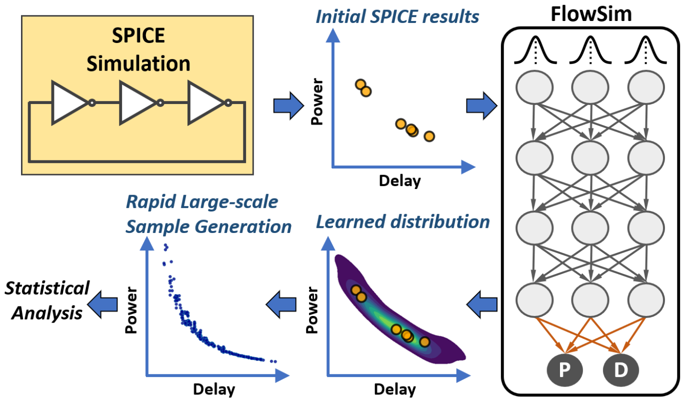


Fig. 1: **The FlowSim framework.**

with a limited number of SPICE simulation results. The primary objective was to approximate the target distribution, represented by an extensive set of Monte Carlo simulations. To evaluate the effectiveness of FlowSim, we compared its performance with two baseline methods: Multivariate Normal (MVN) and Gaussian Kernel Density Estimation (KDE). The MVN baseline was implemented by fitting a multivariate Gaussian distribution to the data, estimating the mean vector and covariance matrix from the provided samples. In the case of KDE, we performed non-parametric density estimation using a Gaussian kernel.

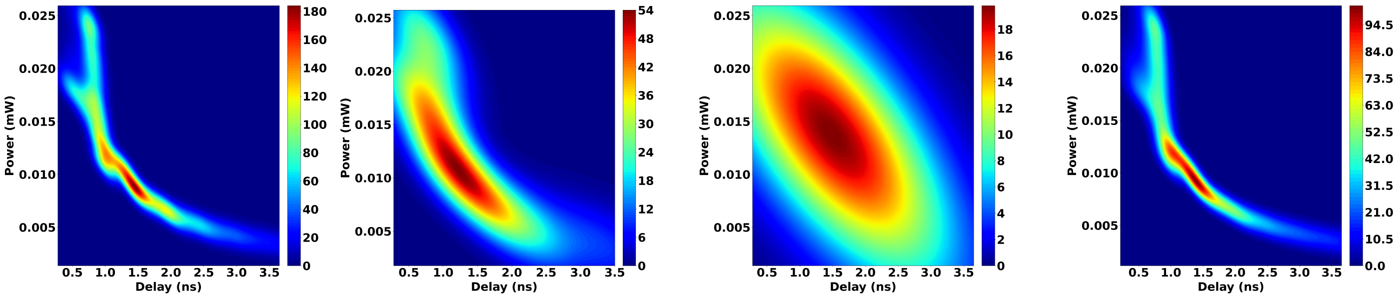
Simulations were performed on ROs designed with a standard CMOS process. We employed the BSIM4 model with various process variables such as t_{ox}, V_{th0}, v_{sat}, μ_0, and N_{factor} to simulate changes in RO behavior due to process variations, as detailed in Table II. A dataset of 10^6 power and delay simulation results with varying V_{DD} was generated. The methods were trained on different numbers of simulations (from 100 to 10,000), and performance was evaluated based on the accuracy of calculated statistics such as mean (μ), standard deviation (σ), sigma values, and yield estimation.

FlowSim was designed with 16 layers, each containing a permutation and a coupling sub-layer. We used a batch size of 64 and a learning rate of 1e-4, which was adjusted based on performance. The loss function was defined as the KL divergence between the flow network and the training data distribution. The training data were split into both training and validation sets. Training epochs were set to be sufficient, and early stopping was implemented to prevent overfitting when the validation loss stopped decreasing. To further prevent overfitting, we included an L2-norm regularization via a weight decay term in the loss function, ensuring the weights in the network did not explode during training.

Fig 3 demonstrates the evolution of 1000 randomly generated samples from a Gaussian distribution as they pass through each layer of the FlowSim model. This model was trained on a 51-stage RO at $V_{DD} = 0.9V$. As the random samples pass through the deeper layers of the network, they gradually begin to resemble the true distribution of the data.

V. Performance Comparison

Figure 4 demonstrates that the kernel density estimation (KDE) and multivariate normal (MVN) methods exhibit a pronounced bias in the probability distribution, leading to an approximation of the target distribution that resembles a Gaussian-like shape. This characteristic presents challenges for these techniques when dealing with complex target distributions. In contrast, FlowSim, due to the significant expressive power of invertible neural networks, can effectively learn arbitrary target densities.



(a) **True Density** (b) **Kernel Density Estimation** (c) **Multivariate Normal** (d) **FlowSim (Ours)**

Fig. 2: **Density estimation results for FoMs (power and delay) of a 5-stage RO with $V_{DD} = 0.7V$.**

TABLE I: **Comparison of methods approximating MC simulations for a 101-stage RO:** Estimated statistics and their error rates relative to reference values. Each method learns from only 100 samples, with our approach showing rapid and accurate convergence to the reference distribution.

Method	Delay (ns)					Power (mW)				
	μ	σ	95.5%	99.7%	99.9%	μ	σ	95.5%	99.7%	99.9%
Reference	2.04	0.19	2.47	2.70	2.87	0.67	0.08	0.85	0.93	0.99
MVN	2.05(0.3%)	0.20(9.5%)	2.54(3.0%)	**2.69(0.3%)**	2.70(5.8%)	0.66(0.3%)	0.08(5.0%)	0.85(0.7%)	0.89(4.3%)	0.89(10.1%)
KDE	2.05(0.3%)	0.23(20.9%)	2.59(5.0%)	2.82(4.3%)	2.95(2.8%)	0.66(0.4%)	0.09(15.8%)	0.87(2.7%)	0.95(1.8%)	1.00(0.9%)
FlowSim	2.05(0.3%)	**0.18(4.6%)**	**2.45(0.9%)**	2.68(0.8%)	**2.86(0.2%)**	0.66(0.3%)	**0.08(2.0%)**	**0.85(0.5%)**	**0.92(0.9%)**	**0.98(0.8%)**

95.5%, 99.7%, and 99.9% values denote right tail quantiles at 2, 3, and 4 sigma levels, computed from 100,000 generated actual samples.

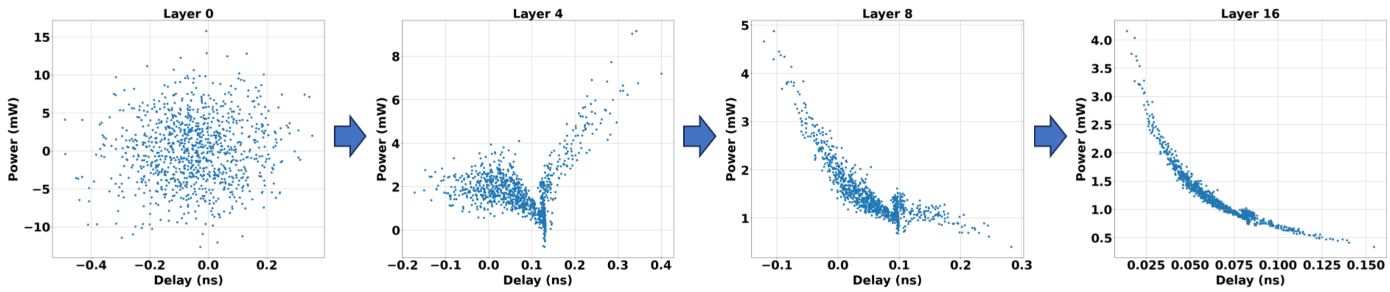


Fig. 3: **Transformation of 1000 Random Samples through FlowSim Layers.** The figure illustrates how the initial random distribution evolves towards the target distribution across the layers of the FlowSim model trained on a 51-stage RO at $V_{DD} = 0.9V$.

TABLE II: **Process variable ranges for NMOS and PMOS.**

Process Variable	**Symbol**	**Unit**	**NMOS Range**	**PMOS Range**
Oxide Thickness	t_{ox}	nm	1.53–2.07	1.46–1.98
Threshold Voltage	V_{th0}	V	0.529–0.716	−0.675–−0.499
Velocity Saturation	v_{sat}	m/s	110,500–149,500	76,500–103,500
Mobility	μ_0	m^2/Vs	0.0417–0.0564	0.0179–0.0242
Subthreshold Swing	N_{factor}	–	1.36–1.84	1.53–2.07

As presented in Table I, the issues related to Gaussian-like approximation become evident. The conventional concept of 2, 3, and 4 sigma values can diverge substantially from the actual quantiles. For instance, the right tail quantiles of the 4 sigma value for power underestimate the true 99.9% quantile value by 10.1%, a discrepancy that could lead to issues during circuit verification. These findings underscore the importance of utilizing advanced statistical methods like FlowSim to accurately learn target distributions, allowing for a deeper understanding of the distribution. Such insights can guide more informed decision-making in circuit design and verification.

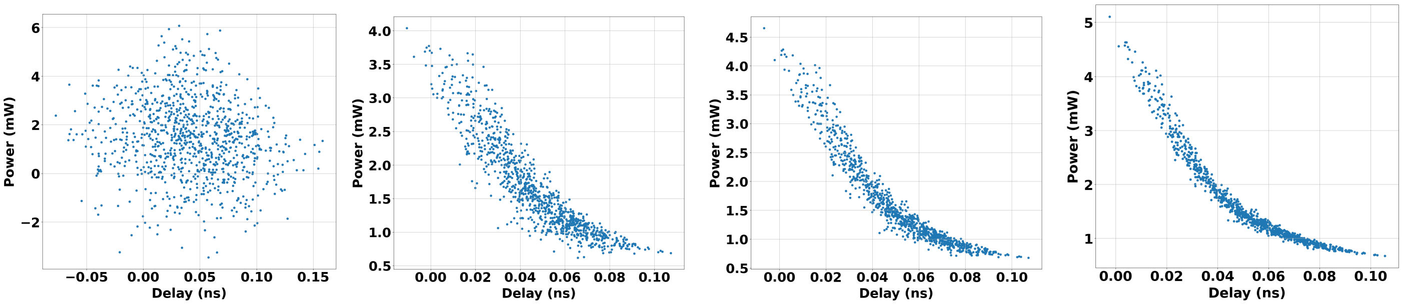
Table III illustrates the yield estimation performance for three distinct sets of required power and delay specifications of the 5-stage RO with V_{DD} at 1.1V. Across all conditions, FlowSim consistently outperforms the other baselines. The error rates of FlowSim, compared to the true baseline, range from as low as 0.2% to as high as 2.8%, demonstrating remarkable accuracy.

It is worth noting that all our experimental findings suggest the limitations of relying solely on summary statistics, such as mean and variance, a notion well-established in the literature. The classic example of Anscombe's quartet [2] effectively illustrates how datasets with identical summary statistics can have vastly different underlying distributions. These examples highlight the necessity of employing advanced statistical techniques, such as FlowSim, to attain a comprehensive understanding of the data and make informed decisions based on this understanding.

TABLE III: **Yield estimation accuracy comparison.**

Method	**D $\leq$ 0.19ns P $\leq$ 0.26mW**	**D $\leq$ 0.20ns P $\leq$ 0.28mW**	**D $\leq$ 0.21ns P $\leq$ 0.29mW**
Reference	70.0	80.0	90.0
MVN	67.1(4.1%)	78.6(1.8%)	89.1(1.0%)
KDE	66.7(4.7%)	77.5(3.1%)	87.5(2.8%)
FlowSim	**68.0(2.8%)**	**79.8(0.2%)**	**89.5(0.5%)**

D and P denote delay and power, respectively.



(a) **Epoch 10** (b) **Epoch 20** (c) **Epoch 30** (d) **Epoch 40**

Fig. 4: **Output Distribution Progression Over Training Epochs.** Using the FlowSim model trained on a 51-stage RO at $V_{DD} = 0.9V$, the figure illustrates how the model's output distribution aligns with the target distribution as training progresses.

VI. Conclusion

This paper presents FlowSim, an invertible neural network designed to learn the statistics of Figure of Merit (FoM) for a circuit using significantly less data than traditional Monte Carlo simulations. This paper is the first to use direct density estimation of circuit FoMs for statistical analysis. The experiments conducted show that FlowSim is capable of accurately estimating target distributions and performing yield and high sigma estimation with over a 100x reduction in the number of simulations. The results are promising and suggest that FlowSim can significantly reduce simulation time and circuit optimization cycles, potentially impacting the field of statistical analysis of process variations.

References

[1] 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.

[2] F. J. Anscombe. Graphs in statistical analysis. *The American Statistician*, 27(1):17–21, 1973.

[3] Wim Bogaerts, Yufei Xing, and Umar Khan. Layout-aware variability analysis, yield prediction, and optimization in photonic integrated circuits. *IEEE Journal of Selected Topics in Quantum Electronics*, 25(5):1–13, 2019.

[4] Tae Hoon Choi, Hanwool Jeong, and Seong-Ook Jung. Fast monte-carlo analysis method of ring oscillators with neural networks. In *2018 International Conference on Electronics, Information, and Communication (ICEIC)*, pages 1–4. IEEE, 2018.

[5] Aksh Chordia, Surendra Hemaram, and Jai Narayan Tripathi. An automated framework for variability analysis using simulated annealing. In *2021 IEEE 25th Workshop on Signal and Power Integrity (SPI)*, pages 1–4. IEEE, 2021.

[6] Wei He, Yigang He, and Bing Li. Generative adversarial networks with comprehensive wavelet feature for fault diagnosis of analog circuits. *IEEE Transactions on Instrumentation and Measurement*, 69(9):6640–6650, 2020.

[7] Ivan Kobyzev, Simon J.D. Prince, and Marcus A. Brubaker. Normalizing flows: An introduction and review of current methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(11):3964–3979, 2021.

[8] Slawomir Koziel and Adrian Bekasiewicz. Variable-fidelity response feature surrogates for accelerated statistical analysis and yield estimation of compact microwave components. *IET Microwaves, Antennas & Propagation*, 13(14):2539–2543, 2019.

[9] Hiwa Mahmoudi and Horst Zimmermann. A new sampling technique for monte carlo-based statistical circuit analysis. In *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2017*, pages 1277–1280. IEEE, 2017.

[10] Siddhartha Nath, Geraldo Pradipta, Corey Hu, Tian Yang, Brucek Khailany, and Haoxing Ren. Generative self-supervised learning for gate sizing. In *Proceedings of the 59th ACM/IEEE Design Automation Conference*, pages 1331–1334, 2022.

[11] 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.

[12] 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.

[13] Shupeng Sun, Xin Li, Hongzhou Liu, Kangsheng Luo, and Ben Gu. Fast statistical analysis of rare circuit failure events via scaled-sigma sampling for high-dimensional variation space. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 34(7):1096–1109, 2015.

[14] Dennis D Weller, Michael Hefenbrock, Michael Beigl, and Mehdi B Tahoori. Fast and efficient high-sigma yield analysis and optimization using kernel density estimation on a bayesian optimized failure rate model. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 41(3):695–708, 2021.

[15] Biying Xu, Yibo Lin, Xiyuan Tang, Shaolan Li, Linxiao Shen, Nan Sun, and David Z Pan. Wellgan: Generative-adversarial-network-guided well generation for analog/mixed-signal circuit layout. In *Proceedings of the 56th Annual Design Automation Conference 2019*, pages 1–6, 2019.

[16] Wei Ye, Mohamed Baker Alawieh, Yibo Lin, and David Z Pan. Lithogan: End-to-end lithography modeling with generative adversarial networks. In *Proceedings of the 56th Annual Design Automation Conference 2019*, pages 1–6, 2019.

[17] Keren Zhu, Mingjie Liu, Yibo Lin, Biying Xu, Shaolan Li, Xiyuan Tang, Nan Sun, and David Z Pan. Geniusroute: A new analog routing paradigm using generative neural network guidance. In *2019 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pages 1–8. IEEE, 2019.