

Reversible Gating Architecture for Rare Failure Detection of Analog and Mixed-Signal Circuits

Myung Seok Shim

Electrical and Computer Engineering
Texas A&M University
College Station, Texas 77843, USA
mrshim1101@tamu.edu

Hanbin Hu

Electrical and Computer Engineering
University of California, Santa Barbara
Santa Barbara, CA 93106, USA
hanbinhu@ucsb.edu

Peng Li

Electrical and Computer Engineering
University of California, Santa Barbara
Santa Barbara, CA 93106, USA
lip@ucsb.edu

Abstract—Due to the growing complexity and numerous manufacturing variation in safety-critical analog and mixed-signal (AMS) circuit design, rare failure detection in the high-dimensional variational space is one of the major challenges in AMS verification. Efficient AMS failure detection is very demanding with limited samples on account of high simulation and manufacturing cost. In this work, we combine a reversible network and a gating architecture to identify essential features from datasets and reduce feature dimension for fast failure detection. While reversible residual networks (RevNets) have been actively studied for its restoration ability from output to input without the loss of information, the gating network facilitates the RevNet to aim at effective dimension reduction. We incorporate the proposed reversible gating architecture into Bayesian optimization (BO) framework to reduce the dimensionality of BO embedding important features clarified by gating fusion weights so that the failure points can be efficiently located. Furthermore, we propose a conditional density estimation of important and non-important features to extract high-dimensional original input features from the low-dimension important features, improving the efficiency of the proposed methods. The improvements of our proposed approach on rare failure detection is demonstrated in AMS data under the high-dimensional process variations.

Index Terms—Reversible neural network, dimension reduction, Bayesian optimization, gating architecture, failure detection

I. INTRODUCTION

Analog and mixed-signal (AMS) systems demand very strict requirement from design to tape-out, especially for the safety-critical applications. For example, in biomedical device, an extremely low failure rate is typically required such as 1 DPPM (defective parts per million) or less, which makes the circuit design and the corresponding circuit verification very challenging tasks.

Nowadays, the most common practice for analog circuit verification in industry is still to use Monte Carlo (MC) simulations to detect rare failures. However, MC methods are computationally expensive in terms of long simulation time with stringent verification requirements. In recent years, as one of data-efficient optimization methods brought by the machine learning society, Bayesian optimization (BO) is introduced into the circuit verification field for failure rate estimation and rare failure detection [1]–[4]. BO is a sequential search mechanism for optimization of black-box objective functions which are expensive for evaluation. In particular, BO trains a surrogate model to represent the objective function, and optimizes an

acquisition function based on the surrogate model to guide the sampling process. [1], [2] present importance sampling method with BO to find the global extreme value and high trustworthy estimated failure detection with few circuit simulations. [3] proposed to apply multiple acquisition functions to guide the sampling process for well-balanced exploitation and exploration to achieve fast rare failure detection.

Typically, AMS circuits usually suffer from a large number of process variations, which makes AMS verification a high-dimensionality problem. However, when it comes with high dimensional space, the training of the surrogate model and the optimization for acquisition function becomes increasingly inefficient, resulting in less effective failure detection.

To resolve the aforementioned issue, a dimension reduction scheme should be introduced into the BO framework. [4] proposed a random embedding technique to reduce the dimensionality of BO for efficient failure detection. However, such embedding provides no indicator for the dimension reduction quality, and only extracts a linear embedding for the high-dimensional data, which cannot handle nonlinear manifold. Instead, we seek out some recent development in the neural network architectures to resolve these two issues. First, [5] proposed a gating architecture called ARGate identifying essential and non-essential features via fusion weights representing feature importance regulated by an auxiliary loss for each feature. The fusion weights can be used as an indicator for the dimension reduction quality. On the other hand, recently developed architectures, reversible residual networks (RevNets) are well known for their representation learning with no information loss [6]–[8], which can be well served to extract nonlinear manifold for the dimension embedding.

In this paper, we propose a RevNet based gating neural network with the improved performance for the rare failure detection problem using the BO framework. Our main contributions are: **1)** propose a new RevNet based auxiliary-model regulated gating architecture, called *Rev-Gate*, to utilize gating fusion weights for efficient dimension reduction; **2)** propose a novel dimension embedding method using RevNet and Bayesian neural network (BNN) to embed low-dimensional nonlinear internal representation back into the high-dimensional original variation parameters; and **3)** investigate the proposed dimension embedding in a BO framework for efficient rare failure detection via extensive experimental studies. We demonstrate in the experimental study that our proposed Rev-Gate archi-

ture efficiently detects rare AMS failures with significantly less runtime while other methods don't.

II. BAYESIAN OPTIMIZATION FOR HIGH DIMENSION

A. Failure Detection Problem Formulation

Under a given D -dimensional parameter space $\mathcal{X} \subseteq \mathbb{R}^D$, the goal of failure detection is to find a failure point \mathbf{x} to meet the following requirement:

$$\exists \mathbf{x} \in \mathcal{X}, y(\mathbf{x}) < T, \quad (1)$$

where T is the threshold target for specification requirement, and the $y(\mathbf{x})$ represents the performance of circuit at the parameter variation vector \mathbf{x} . When the value of $y(\mathbf{x})$ is smaller than the threshold T , the performance is considered as the failure with the specific point \mathbf{x} . Due to the nature of $y(\mathbf{x})$ which is severely nonlinear in the high dimensional space, it is hard and costly to get this value in terms of simulation time and computational resources. Instead, we reformulate the previous failure detection problem into an optimization issue below to fit into the Bayesian optimization context.

$$\min_{\mathbf{x} \in \mathcal{X}} y(\mathbf{x}) < T \quad (2)$$

B. High Dimensional Bayesian Optimization

Bayesian optimization incorporates two major modules: a surrogate model and an acquisition function. The surrogate model serves as an approximation for the original black-box function under the optimization, while providing the uncertainty estimation for the current model. One of the most popular implementation for the surrogate model is Gaussian process (GP) defining a normal posterior probability model as $y|\mathbf{x}, \mathcal{D} \sim \mathcal{N}(\mu(\mathbf{x}), \sigma^2(\mathbf{x}))$, where \mathcal{D} is the dataset or observations, $\mu(\mathbf{x})$ and $\sigma^2(\mathbf{x})$ are posterior mean and variance estimation, respectively. In order to efficiently guide the search process to the optimal location, based on the surrogate model, an acquisition function $\alpha(\mathbf{x}|\mathcal{D})$ is optimized to balance the exploitation over optimal solution and the exploration for uncertain areas.

However, the traditional BO suffers from costly GP training and poor optimization quality of the acquisition function over high-dimensional space, which leads to inefficient black-box optimization. One way to mitigate this effect is to embed the original high dimensional space $\mathcal{X} \subseteq \mathbb{R}^D$ into a low dimension space $\mathcal{Z} \subseteq \mathbb{R}^d$, where $d < D$, so that both surrogate model and acquisition function can be performed in a low dimensional space \mathcal{Z} for fast training convergence and better acquisition function optimization. After optimized \mathbf{z}^* is extracted from the acquisition function, it can be embedded back to the original space \mathcal{X} for the actual circuit simulation as shown in Fig. 1 via a dimension embedding process $\mathbf{x}^* = E_{\mathcal{Z} \rightarrow \mathcal{X}}(\mathbf{z}^*)$.

One well-known dimensional embedding method for Bayesian optimization is random embedding [4], [9]. The key idea is to sample a random matrix $A \in \mathbb{R}^{D \times d}$ with each matrix element following a distribution of $\mathcal{N}(0, 1)$ and fix it during the BO process. Therefore, dimension embedding function can be defined as $\mathbf{x}^* = E_{\mathcal{Z} \rightarrow \mathcal{X}}(\mathbf{z}^*) = A\mathbf{z}^*$ to serve as a linear

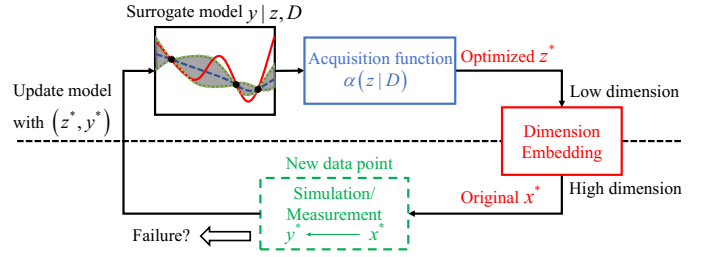


Fig. 1: Bayesian optimization for high-dimensional problems. dimension reduction method for enhanced BO performance under high dimensional space.

Even though the random embedding method reduces the BO dimension effectively, there still exist several major concerns for this method. Firstly, since random embedding is agnostic to the black-box function under optimization, it is hard to decide the low dimension d for random matrix generation beforehand. In addition, the dimension embedding quality is unknown before the actual BO process. Secondly, the random embedding only performs the dimension embedding in a linear manner, which cannot be utilized when a non-linear low dimensional manifold is desired. Finally, it provides no information about the actual important variational parameters, which is essential in aspect of failure detection field for circuit designers to gain more insights about the circuit behavior.

III. PROPOSED REV-GATE BASED BAYESIAN OPTIMIZATION

To tackle the challenges introduced by random embedding in BO for high-dimensional failure detection as mentioned in the previous section, we propose a Rev-Gate architecture for the dimension embedding in the BO framework, which incorporates the RevNet and the ARGate to effectively identify important variational parameters and reduce the dimension through the reversibility. In order to learn the low-dimensional manifold property from the black-box function under optimization, we pre-train the Rev-Gate architecture before the BO process by using a small amount of data, which extracts the important feature information and helps choose effective low dimension d for surrogate model construction and acquisition function optimization. During the BO process, the trained RevNet performs the dimension embedding in Fig. 1, recovering the low dimensional point \mathbf{z}^* to the original input space \mathbf{x}^* for actual circuit simulation via a restoration scheme, which will be further discussed in Section IV. The rest of this section mainly talks about how to identify important features for BO dimension embedding via Rev-Gate pre-training.

A. ARGate for important feature extraction

For AMS failure detection under high-dimensional space, typically there exists certain redundancy for the variational parameters under consideration, and only a small number of them are critical to the final circuit performance. With only important features utilized and inessential ones removed, the circuit performance can still be predicted nicely via the surrogate model even with small amount of training data. To efficiently identify the important variational parameters, we

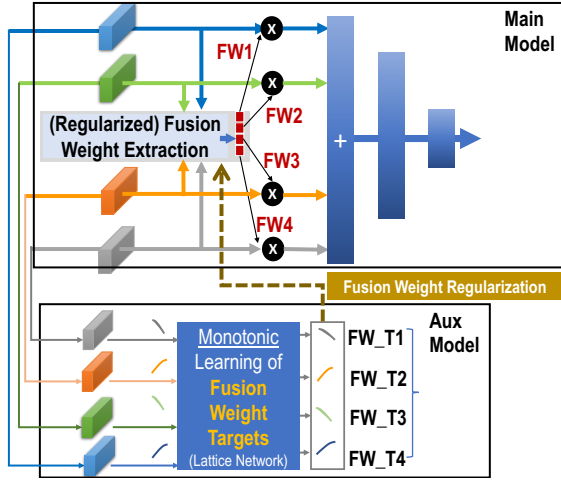


Fig. 2: The ARGate architecture overview.

adopt the ARGate [5] using gating architecture to switch off inessential feature via fusion weights.

In terms of network structure, the ARGate is composed of two networks as shown in Fig. 2: a main model and an auxiliary (aux) model. The fusion weights are extracted from the grey box (denoted as “Fusion Weight Extraction” in Fig. 2) in the main model, where the fusion happens with pre-processed features after a fully connected (FC) layer in each feature path.

The key idea of the ARGate is that the importance of features is represented via the fusion weights. In the main model, the output of each FC layer on each feature path are multiplied with the corresponding fusion weights to obtain a weighted internal representation, which is then passed to later network layers to get final classification/regression output. These weights are normalized between $[0, 1]$ for the feature importance interpretability. For example, assume that there are only four features under consideration. If the first feature is the only important feature in the datasets, the corresponding first fusion weight $FW1$ is the largest fusion weight out of four, which is close to 1. Then, then the fusion weights of other features $FW2$, $FW3$, and $FW4$ are relatively close to 0. As the multiplication mechanism of the fusion weights which switches off the unimportant feature path, the first feature makes a larger impact on the target prediction than the other features.

The auxiliary model is added here to facilitate the reliable training for the fusion weights, regularizing the fusion weights with auxiliary losses reflecting the relevance between the target value and each individual feature.

B. Bijective RevNet for Non-Linear Representation Learning

The ARGate identifies the important input features through fusion weights, which serves as a great tool for dimensionality reduction. However, if we directly apply ARGate on the raw variation parameters, the reduced dimension is only a subset of original variation parameters, which completely ignores the correlation between different variational parameters. Therefore, we applied reversible residual network (RevNet) [8], [10], [11] to learn the correlation between multiple features, and

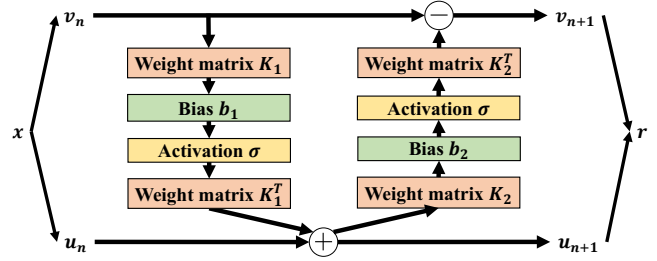


Fig. 3: The reversible block.

forms a non-linear internal representation for original features. In addition, one of major advantage of using RevNet is that it avoids information loss between its input and output, giving a bijective function. Fig. 3 gives a typical RevNet block with the feature mixing process given follows.

$$\begin{aligned} u_{n+1} &= u_n + hK_{n,1}^T \sigma(K_{n,1}v_n + b_{n,1}), \\ v_{n+1} &= v_n - hK_{n,2}^T \sigma(K_{n,2}u_{n+1} + b_{n,2}), \end{aligned} \quad (3)$$

where n ranges from $[0, N - 1]$ for a RevNet with N RevNet blocks, u_n and v_n are two partitions for the n th state with same dimensionality, and h is a scaling factor.

Here we denote the RevNet nonlinear representation learning with $g : \mathbf{x} \mapsto \mathbf{r}$, which goes through N blocks of (3) as follows.

$$\mathbf{x} = \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} \xrightarrow[g^{-1}]{} \begin{bmatrix} u_N \\ v_N \end{bmatrix} = \mathbf{r}, \quad (4)$$

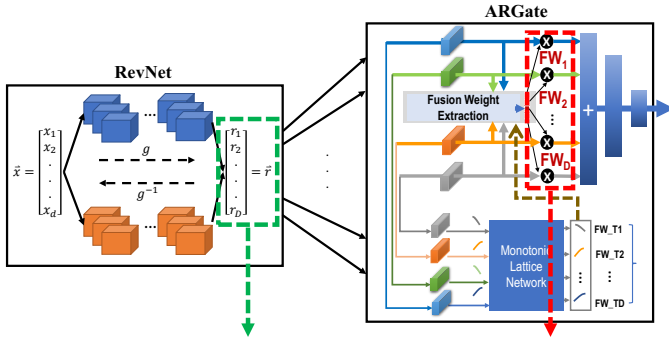
where $u_0 := (x_1, \dots, x_{[D/2]})^T$ and $v_0 := (x_{[D/2]+1}, \dots, x_D)^T$ are the two partitions of the input vector \mathbf{x} , and $u_N := (r_1, \dots, r_{[D/2]})^T$ and $v_N := (r_{[D/2]+1}, \dots, r_D)^T$ are the two partitions of the RevNet output \mathbf{r} .

C. Proposed Rev-Gate architecture

The proposed Rev-Gate architecture connects a RevNet and an ARGate in serial for efficient dimension reduction. In order to utilize such architecture for high-dimensional Bayesian optimization, we first pre-train the proposed architecture using a small amount of data to identify good dimension size d for dimension reduction, and then we utilize the trained RevNet in the reverse direction to embed the low dimension optimized \mathbf{z}^* from acquisition function into \mathbf{x}^* in the original high-dimension space.

1) Dimension Reduction via the Proposed Architecture:

Thanks to the nonlinear bijective characteristic of the RevNet, we generate an internal representation $\mathbf{r} = g(\mathbf{x})$ mapped from the original variation parameter \mathbf{x} , which share the same dimensionality D as \mathbf{x} . With the feature importance interpretability from the ARGate, the importance of each internal representation dimension r_i can be estimated via the corresponding trained fusion weight FW_i . Given user-defined importance threshold FW_{TH} , the dimensionality d of low dimension space \mathcal{Z} in BO can be determined by the number of fusion weights larger than FW_{TH} . The corresponding d internal representation elements can be reassembled into the low dimension feature $\mathbf{z} = (r_{i_1}, \dots, r_{i_d})^T$ used in BO, with each element r_{i_j} having $FW_{i_j} \geq FW_{TH}$, achieving the dimension reduction for surrogate model and acquisition



Identification of important representation features via fusion weights

Fig. 4: The proposed Rev-Gate architecture. Important internal representation features are identified via fusion weight values.

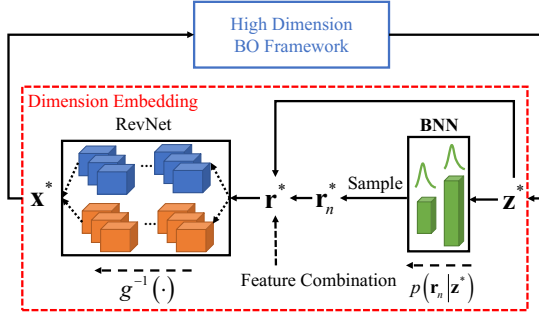


Fig. 5: The proposed BO dimension embedding using RevNet and BNN.

function. The rest $D - d$ elements in \mathbf{r} are considered as noncritical, and marked as \mathbf{r}_n here.

2) *Dimension Embedding using Reverse RevNet*: As shown in Fig. 1, the BO framework provides a \mathbf{z}^* with a dimensionality of d during each iteration, which needs to be embedded into the original space \mathcal{X} . Here we use the trained RevNet in a reverse direction to map low-dimensional \mathbf{z}^* back to \mathbf{x}^* in the high dimensional space. However, to use RevNet for the restoration of the original variational parameters \mathbf{x}^* , it requires the same dimensionality for the input and the output of RevNet. Therefore, $D - d$ new elements should be generated and combined with \mathbf{z}^* to obtain the restored internal representation \mathbf{r}^* . More details about this conversion from \mathbf{z}^* to \mathbf{r}^* is discussed in Section IV. Given the reversibility of RevNet without information loss, we can easily recover the original variation parameters using $\mathbf{x}^* = g^{-1}(\mathbf{r}^*)$ to perform the required dimension embedding in BO as shown in Fig. 5.

IV. ENHANCED DIMENSION EMBEDDING VIA BAYESIAN NEURAL NETWORK

As mentioned in Section III-C2, additional elements should be appended to \mathbf{z}^* to ensure the resulting \mathbf{r}^* sharing the same dimensionality as \mathbf{x}^* for traversing the RevNet in the reverse direction. As we know from the fusion weights, the additional elements are less critical for the final circuit performance. Hence, the simplest approach here is to append zeros to \mathbf{z}^* to the high dimensionality D .

However, the zero appending approach neglects the correlation between \mathbf{z}^* and \mathbf{r}_n which both depend on the original input vector \mathbf{x} . Instead, we propose to learn a conditional

probability distribution $p(\mathbf{r}_n | \mathbf{z}^*)$ to recover \mathbf{r}^* from \mathbf{z}^* . Here, the particular probabilistic model we used for this conditional distribution is a Bayesian neural network (BNN).

After the Rev-Gate is trained, with the fixed RevNet, we can generate the internal representation \mathbf{r} , and separate them into important features \mathbf{z} and non-important ones \mathbf{r}_n for all the training data. Then the conditional distribution $p(\mathbf{r}_n | \mathbf{z}^*)$ represented by the BNN is estimated with maximum likelihood estimation using these pre-processed data. During dimension embedding in the BO process, a new \mathbf{r}_n^* is randomly sampled from the learnt $p(\mathbf{r}_n | \mathbf{z}^*)$ using a trained BNN, and then combined with \mathbf{z}^* to obtain recovered internal representation \mathbf{r}^* for the RevNet conversion. The complete dimension embedding illustration is presented in Fig. 5.

V. EXPERIMENTAL RESULTS

A. Experimental Setups

We demonstrated our proposed Rev-Gate architecture with BO approach with two circuits: a low-dropout (LDO) regulator [12] (60 dimensions) and a DC-DC converter [13] (44 dimensions), as shown in Fig. 6 and 7.

For the rare failure detection performance comparison, we compared our proposed architectures with Monte Carlo (MC), expected improvement (EI), probability of improvement (PI) in [14], parallelizable Bayesian optimization (pBO) in [3] and parallelizable Bayesian optimization with random embedding (HDBO) in [4]. The BayesOpt [15] was utilized for implementing BO methods. All the simulations were run on a workstation with a 3.50GHz Intel(R) Xeon(R) E5-1620 v4 CPU.

The proposed Rev-Gate is implemented with Pytorch 1.2 [16]. To be specific with the training process, the Rev-Gate is pre-trained with a small amount of the circuit simulation samples which are uniformly distributed in a pre-defined hyper-cube space. For a fair comparison with different BO based methods, we matched total simulation budget for all the BO methods including the number of training samples for the Rev-Gate. After the training phase, fusion weight values were examined for top- d indexes of important features extraction and screening out some non-important features. With the indexes of the important features, a simple BNN is trained for the non-essential component conditional distribution estimation. Finally, with the trained RevNet and the BNN, the BO framework is operated so that the \mathbf{z}^* vector is computed with BNN and RevNet in reverse direction to generate \mathbf{x}^* . The \mathbf{x}^* is used as input sample for the circuit simulation and the simulation output \mathbf{y}^* is passed onto the BO surrogate model.

1) *Low-dropout Regulator*: Three specifications, quiescent current, undershoot and load regulation, are chosen as the verification targets of the LDO regulator. Three kinds of transistor-level variations are considered for all 20 transistors: channel length, threshold voltage and gate oxide thickness, resulting in a 60 dimension variation space. 330,000 samples are used for MC without a single failure detected, suggesting the failures are extremely rare for the LDO circuit. The detailed simulation budget for each method is listed in Table I.

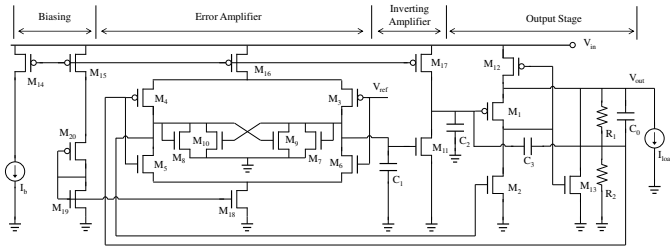


Fig. 6: A low-dropout regulator.

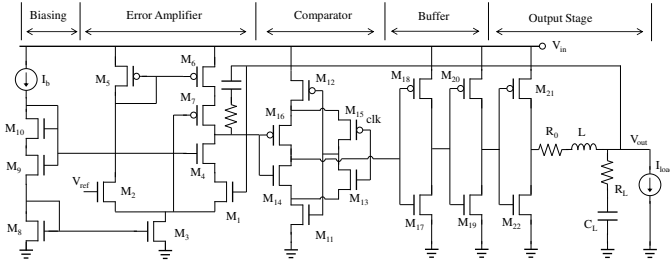


Fig. 7: A pwm/pfm dc-dc converter.

Note that the actual simulation for PI and pBO for undershoot and load regulation is slightly less than other ones, as shown in Table I, due to the accidental program break. However, we can see that the corresponding runtime is still much larger than the proposed methods, which makes them still inefficient for failure detection under high dimensional space.

The number of essential features extracted from the proposed architecture is 26 out of 60 for quiescent current and load regulation, 30 for undershoot. The same dimension reduction is used for HDBO. Furthermore, in the light of circuit aspects, the Rev-Gate identify the actual important features from the inputs while HDBO cannot. We observe that most important parameters are located on the output stage in the LDO regulator, which is close to the circuit designers' insights.

2) *DC-DC converter*: As shown in Fig. 7, total 22 transistors are included in the DC-DC converter with two variational parameters for each transistor: channel length and width, resulting in 44 input features for the simulation. Two specifications are considered: output accuracy and overshoot. Through our proposed Rev-Gate, we could reduce the number of dimension from 44 to 14 for output accuracy, and 16 for overshoot, which is far less than half of the total number of input features. Detailed simulation budget setup is included in Table II.

B. Failure Detection Results

From Table I and II, MC, EI, PI, pBO and HDBO methods cannot detect a failure case due to the challenging rare failure detection in the high-dimensional parameter space. On the other hand, our proposed Rev-Gate based BO framework successfully find the worst case for all specifications with in the simulation budget thanks to proposed Rev-Gate architecture for dimension embedding in BO. In terms of the magnitude of the worst case detected, MC, EI, PI typically cannot find any worse case near the target for most specification under

consideration. pBO and HDBO presents a better performance with its good exploration and exploitation balancing, while the proposed architecture presents the worst case detected than all other methods for all the specifications with the help of effective dimension embedding given by the Rev-Gate architecture.

Regarding simulation running time, overall BO based methods like EI, PI and pBO take much longer than our proposed Rev-Gate with BNN due to high overhead introduced by surrogate model and acquisition function in high dimension. HDBO suffers from its simple dimension embedding mechanism to achieve poor failure detection efficiency. With a smart sampling budget allocation for Rev-Gate pre-trained dimension reduction, the BO search efficiency is significantly improved leading to short runtime. Note that the runtime for Rev-Gate with BNN includes pre-training phase.

C. Worst Case Trend Analysis

Finally, the worst case trend is analyzed as shown in Fig.8. BO based methods such as EI and PI found the worst case slowly comparing to pBO and HDBO. During the first 500 samples, EI, PI, and pBO shows similar failure detection performance but the worst case of pBO rises after the 500 samples. The worst case of HDBO was bit larger than other BO based methods but it is stuck at local minima around 600 samples. Our proposed Rev-Gate shown in green color in the graph starts with the lowest worst case, it found its worst case much more rapidly than the other methods at the initial search process. Note that 500 samples are used for pre-training Rev-Gate architecture. From this result, it is clear that pre-training process shows significant benefits for the improved failure detection performance and efficiency.

VI. CONCLUSION

In this paper, we present the Rev-Gate architecture with a novel restoration scheme via Bayesian neural network. The proposed algorithm works under Bayesian optimization for rare failure detection of analog mixed-signal circuits. The ARGate is adopted for the identification of important features and the RevNet is utilized for input restoration via backward

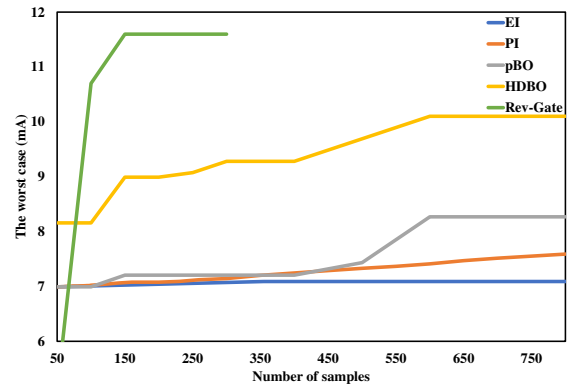


Fig. 8: A plot of the worst case found in quiescent current in the LDO regulator. Note that Rev-Gate with BNN only runs 300 samples on the BO framework for fair comparison.

TABLE I: Failure detection result comparison for the LDO regulator (60 dimension).

Spec	Target	Method	# Sim	Worst Case	1st Failure Hit	Runtime
Quiescent Current	11.0mA	MC	330,000	10.8mA	-	47h45m
		EI	$50_{init} + 750_{seq}$	7.1mA	-	24h06m
		PI	$50_{init} + 750_{seq}$	8.3mA	-	23h43m
		pBO	$50_{init} + 5 \times 150_{batch}$	10.5mA	-	23h53m
		HDBO	$50_{init} + 5 \times 150_{batch}$	10.1mA	-	19h36m
		Rev-Gate with BNN	$500_{training} + 50_{init} + 5 \times 50_{batch}$	11.6mA	621	3h32m
Undershoot	0.52V	MC	330,000	0.32V	-	47h45m
		EI	$50_{init} + 1250_{seq}$	0.27V	-	80h25m
		PI	$50_{init} + 1050_{seq}$	0.23V	-	65h48m
		pBO	$50_{init} + 5 \times 140_{batch}$	0.49V	-	19h52m
		HDBO	$50_{init} + 5 \times 250_{batch}$	0.51V	-	32h25m
		Rev-Gate with BNN	$1000_{training} + 50_{init} + 5 \times 50_{batch}$	0.53V	1248	3h29m
Load regulation	58.2%	MC	330,000	36.0%	-	47h45m
		EI	$50_{init} + 1250_{seq}$	28.9%	-	104h39m
		PI	$50_{init} + 1050_{seq}$	13.9%	-	65h04m
		pBO	$50_{init} + 5 \times 140_{batch}$	58.1%	-	19h47m
		HDBO	$50_{init} + 5 \times 250_{batch}$	58.1%	-	32h19m
		Rev-Gate with BNN	$1000_{training} + 50_{init} + 5 \times 50_{batch}$	58.4%	1241	3h41m

TABLE II: Failure detection result comparison for the DC-DC converter (44 dimension).

Spec	Target	Method	# Sim	Worst Case	1st Failure Hit	Runtime
Output accuracy	58mV	MC	40,800	42.4mV	-	47h50m
		EI	$50_{init} + 1250_{seq}$	25.8mV	-	92h44m
		PI	$50_{init} + 1250_{seq}$	22.4mV	-	92h13m
		pBO	$50_{init} + 5 \times 250_{batch}$	57.3mV	-	97h08m
		HDBO	$50_{init} + 5 \times 250_{batch}$	57.7mV	-	21h21m
		Rev-Gate with BNN	$1000_{training} + 50_{init} + 5 \times 50_{batch}$	58.1mV	1177	4h53m
Overshoot	8.8mV	MC	40,800	8.29mV	-	47h50m
		EI	$50_{init} + 1250_{seq}$	7.30mV	-	92h31m
		PI	$50_{init} + 1250_{seq}$	7.37mV	-	92h34m
		pBO	$50_{init} + 5 \times 250_{batch}$	8.65mV	-	97h46m
		HDBO	$50_{init} + 5 \times 250_{batch}$	8.77mV	-	25h21m
		Rev-Gate with BNN	$1000_{training} + 50_{init} + 5 \times 50_{batch}$	8.84mV	1221	3h59m

computation without loss of information. The Bayesian neural network is applied for non-essential parameter estimation under the nature of conditional probability distribution given essential variational inputs. The experimental results show that our proposed algorithm detects rare failure cases in high dimensional space with less amount of time, while Bayesian optimization with traditional and improved acquisition function does not find anomaly during the circuit simulation.

ACKNOWLEDGEMENT

This material is based upon work supported by the National Science Foundation (NSF) under Grant No. 1956313 and Semiconductor Research Corporation (SRC) Task No. 2810.031 through UT Dallas' Texas Analog Center of Excellence. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF and SRC.

REFERENCES

- [1] D. D. Weller, M. Hefenbrock, M. S. Golanbari, M. Beigl, and M. B. Tahoori, "Bayesian optimized importance sampling for high sigma failure rate estimation," in *2019 Design, Automation Test in Europe Conference Exhibition (DATE)*, 2019, pp. 1667–1672.
- [2] M. Hefenbrock, D. D. Weller, M. Beigl, and M. B. Tahoori, "Fast and accurate high-sigma failure rate estimation through extended bayesian optimized importance sampling," in *2020 Design, Automation Test in Europe Conference Exhibition (DATE)*, 2020, pp. 103–108.
- [3] H. Hu, P. Li, and J. Z. Huang, "Parallelizable bayesian optimization for analog and mixed-signal rare failure detection with high coverage," in *Proceedings of the International Conference on Computer-Aided Design*, 2018, pp. 1–8.
- [4] H. Hu, P. Li, and J. Z. Huang, "Enabling high-dimensional bayesian optimization for efficient failure detection of analog and mixed-signal circuits," in *2019 56th ACM/IEEE Design Automation Conference (DAC)*, 2019, pp. 1–6.
- [5] M. S. Shim, C. Zhao, Y. Li, X. Zhang, and P. Li, "Robust deep multi-modal sensor fusion using fusion weight regularization and target learning," *CoRR*, vol. abs/1901.10610, 2019. [Online]. Available: <http://arxiv.org/abs/1901.10610>
- [6] J.-H. Jacobsen, A. Smeulders, and E. Oyallon, "i-revnet: Deep invertible networks," *arXiv preprint arXiv:1802.07088*, 2018.
- [7] A. N. Gomez, M. Ren, R. Urtasun, and R. B. Grosse, "The reversible residual network: Backpropagation without storing activations," in *Advances in neural information processing systems*, 2017, pp. 2214–2224.
- [8] G. Zhang, J. Zhang, and J. Hinkle, "Learning nonlinear level sets for dimensionality reduction in function approximation," in *Advances in Neural Information Processing Systems*, 2019, pp. 13 220–13 229.
- [9] Z. Wang, M. Zoghi, F. Hutter, D. Matheson, N. De Freitas *et al.*, "Bayesian optimization in high dimensions via random embeddings," in *IJCAI*, 2013, pp. 1778–1784.
- [10] B. Chang, L. Meng, E. Haber, L. Ruthotto, D. Begert, and E. Holtham, "Reversible architectures for arbitrarily deep residual neural networks," *arXiv preprint arXiv:1709.03698*, 2017.
- [11] E. Haber and L. Ruthotto, "Stable architectures for deep neural networks," *Inverse Problems*, vol. 34, no. 1, p. 014004, 2017.
- [12] S. Lai and P. Li, "A fully on-chip area-efficient cmos low-dropout regulator with fast load regulation," *Analog Integrated Circuits and Signal Processing*, vol. 72, no. 2, pp. 433–450, 2012.
- [13] Y. Wang, P. Li, and S. Lai, "A unifying and robust method for efficient envelope-following simulation of pwm/pfm dc-dc converters," in *2014 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*. IEEE, 2014, pp. 618–625.
- [14] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, "Taking the human out of the loop: A review of bayesian optimization," *Proceedings of the IEEE*, vol. 104, no. 1, pp. 148–175, 2016.
- [15] R. Martinez-Cantin, "Bayesopt: A bayesian optimization library for nonlinear optimization, experimental design and bandits," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 3735–3739, 2014.
- [16] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, 2019, pp. 8024–8035.