Enabling High-Dimensional Bayesian Optimization for Efficient Failure Detection of Analog and Mixed-Signal Circuits

Hanbin Hu
Department of ECE
Texas A&M University
College Station, Texas
hanbinhu@tamu.edu

Peng Li
Department of ECE
Texas A&M University
College Station, Texas
pli@tamu.edu

Jianhua Z. Huang
Department of Statistics
Texas A&M University
College Station, Texas
jianhua@stat.tamu.edu

ABSTRACT

With increasing design complexity and stringent robustness requirements in application such as automotive electronics, analog and mixed-signal (AMS) verification becomes a key bottleneck. Rare failure detection in a high-dimensional parameter space using minimal expensive simulation data is a major challenge. We address this challenge under a Bayesian learning framework using Bayesian optimization (BO). We formulate the failure detection as a BO problem where a chosen acquisition function is optimized to select the next (set of) optimal simulation sampling point(s) such that rare failures may be detected using a small amount of data. While providing an attractive black-box solution to design verification, in practice BO is limited in its ability in dealing with high-dimensional problems. We propose to use random embedding to effectively reduce the dimensionality of a given verification problem to improve both the quality of BO-based optimal sampling and computational efficiency. We demonstrate the success of the proposed approach on detecting rare design failures under high-dimensional process variations which are completely missed by competitive smart sampling and BO techniques without dimension reduction.

1 INTRODUCTION

With increasing design complexity and stringent robustness requirements, analog and mixed-signal (AMS) verification becomes a key bottleneck [2]. Many AMS verification methods, e.g. smart sampling [7, 18, 19] and statistical blockade [15] were proposed to speed up the verification process in the past decade. Recent years have witnessed an accelerated integration of AMS ICs into safety-critical applications such as auto-electronics and bio-medical systems which may impose a stringent failure rate specification of 1 DPPM (defective parts per million) or less on AMS ICs. Detecting even a single failure for circuits that are designed to be extremely robust with typical simulation budgets during design time is a completely nontrivial problem. Under this case, the rare failure detection problem (e.g. finding the first failure) is a more fundamental and challenging problem than yield prediction that has been focused on in prior work [7, 15, 18, 19].

The key problem this paper aims to address is the rare failure detection of AMS circuits under large number of design uncertainties (e.g. process variations) with a practically constrained simulation data budget. Towards this end, the recent work of [5] adopted Bayesian optimization (BO) for rare failure detection, demonstrating verification of AMS circuits with relatively limited numbers of process parameters. BO is a powerful tool to find the optimum value of black-box functions [14], which is popular for hyper-parameter tuning [16, 20] and reinforcement learning [1] in machine learning. BO is a sequential design technique for global optimization of black-box objective functions that are expensive to evaluate. A chosen acquisition function is optimized at each step to select the next (set of) optimal sampling location(s). Queries of the objective function to be optimized, e.g. performance of an AMS circuit, which can be costly, e.g. via circuit simulations for AMS verification, are only made at these optimized locations. The new data collected at each step augments the training dataset to retrain a probabilistic surrogate model, e.g. Gaussian process model (GP), that approximates the black-box function. The solutions obtained from optimizing the acquisition function determine where the additional data will be sampled, contribute directly to the accuracy of the surrogate model, and guide the iterative global optimization process.

While providing an attractive black-box solution applicable to AMS verification, a well-known limitation of Bayesian optimization is its limitation in dealing with high-dimensional problems [14, 17]. When the dimensionality of the black-box optimization problem increases, so does the dimensionality of the optimization of the acquisition function, which is typically non-convex, at each sequential sampling step. Solving high-dimensional optimization problems can be both computationally expensive and hard. The high run-time cost and degradation of optimization solution quality for high-dimensional problems severely limit the scalability of BO.

This work aims to extend the applicability of BO to the challenging problem of rare failure detection of AMS circuits with large numbers of design uncertainties. We propose to employ *random embedding* [21] to effectively reduce the effective dimensionality of the verification problem. Dimensionality reduction is possible for AMS circuits since under many practical situations variational parameters of a circuit do not have equal significance to a given design performance to be verified [9, 10]. Specific circuit topologies employed in practical circuits build constrained structures into the way different circuit/process parameters interact with each other and influence the given design performance. This gives rise to parameters that are statistically insignificant to the targeted performance. It shall be noted, however, such parametric redundancy in practice may be only identified in a transformed parameter space.

DAC'19, June 2019, Las Vegas, NV USA

ACM ISBN 978-1-4503-6725-7/19/06…$15.00
https://doi.org/10.1145/3316781.3317818

Towards this end, random embedding provides a systematic way to explore hidden parametric redundancy. As such, parameter redundancy needs not to be specified by the designer *a prior*, which is very hard in general. Instead, it can be streamlined in the sequential statistical learning/black-box optimization framework of Bayesian optimization. Our random embedding based BO approach is further supported by a proposed random embedding dimension selection algorithm that estimates the effective dimension of a given AMS circuit using a small amount of training (simulation) data prior to the BO-based failure discovery process. We demonstrate the success of the proposed approach on detecting rare design failures under up to 60 process parameters which are completely missed by competitive smart sampling and BO techniques without dimension reduction.

2 BAYESIAN OPTIMIZATION PRELIMINARY

2.1 Failure Detection Problem Formulation

Given a D-dimensional parameter variational space $\Omega \subseteq \mathbb{R}^D$, failure detection attempts to find the existence of points not satisfying a given specification inside Ω. Without loss of generality, a point $\boldsymbol{x}$ is regarded a failure if:

$$y(\boldsymbol{x}) < T, \boldsymbol{x} \in \Omega, \tag{1}$$

where T is the targeted specification (assuming the smaller the value is, the worse the performance is), and the $y(\boldsymbol{x})$ represents the circuit performance at the parameter variation combination $\boldsymbol{x}$. Typically, the circuit performance has no closed-form expression, and is highly-nonlinear and complex. Instead of solving a SAT problem with the black-box function $y(\boldsymbol{x})$, the failure detection problem can be formulated as an equivalent optimization problem:

$$\min_{\boldsymbol{x} \in \Omega} y(\boldsymbol{x}) < T. \tag{2}$$

We adopt Bayesian Optimization to optimize $y(\boldsymbol{x})$ as a black-box objective function.

2.2 Bayesian Optimization Introduction

There exist two critical components constituting Bayesian optimization, as shown in Fig. 1. The first one is a surrogate probabilistic model $y^* \mid \boldsymbol{x}^*, \mathcal{D}$ to approximate the original optimization objective function. The probabilistic surrogate model comes with model prediction uncertainty, which can be reduced with sequentially collected examples. The new examples are collected via optimizing the other critical component, an acquisition function $\alpha(\boldsymbol{x}; \mathcal{D})$, which is based on the current information provided by the surrogate model. By carefully designing the acquisition function, the search process can be guided either to improve the surrogate model accuracy or to find more optimal objective function values. Queries of the objective function to be optimized are only made at these optimized locations. Hence, Bayesian optimization greatly reduces the number of objective function evaluations, and is well-suited for AMS rare failure detection.

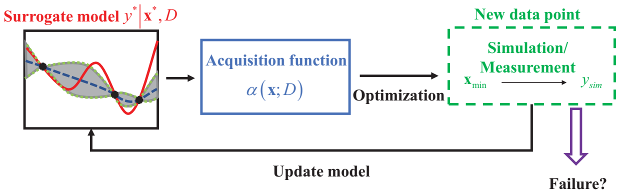


Figure 1: Bayesian optimization procedure.

2.2.1 Surrogate Model. To obtain both mean and uncertainty prediction for the black-box objective function, a Gaussian Process (GP) model is usually chosen as the surrogate model. The GP is characterized with one prior mean function $m(\boldsymbol{x})$ and one prior covariance function $\kappa(\boldsymbol{x}, \boldsymbol{x}')$ as a generative model:

$$\begin{aligned} f &\sim \mathcal{GP}(m, \kappa), & (3)\\ y \mid f, \sigma_0^2 &\sim \mathcal{N}\left(f, \sigma_0^2\right), & (4) \end{aligned}$$

where σ_0^2 is another prior information representing the intrinsic noise variance. Typically, $m(\boldsymbol{x})$ is set be 0, and common choices for the covariance function $\kappa(\boldsymbol{x}, \boldsymbol{x}')$ are the squared exponential (SE) kernel and Matérn kernel [13]. Provided a finite collection of n examples $\mathcal{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$, and a new test point $\boldsymbol{x}^*$, the posterior mean and variance prediction of the objective function at $\boldsymbol{x}^*$ given by the GP model are [13]:

$$\begin{aligned} y^* \mid \boldsymbol{x}^*, \mathcal{D} &\sim \mathcal{N}\left(\mu(\boldsymbol{x}^*; \mathcal{D}), \sigma^2(\boldsymbol{x}^*; \mathcal{D})\right) & (5)\\ \mu(\boldsymbol{x}^*; \mathcal{D}) &= \boldsymbol{k}^{\mathrm{T}}(\boldsymbol{x}^*)\left(\boldsymbol{K} + \sigma_0^2 \boldsymbol{I}\right)^{-1} \boldsymbol{y} & (6)\\ \sigma^2(\boldsymbol{x}^*; \mathcal{D}) &= \kappa(\boldsymbol{x}^*, \boldsymbol{x}^*) - \boldsymbol{k}^{\mathrm{T}}(\boldsymbol{x}^*)\left(\boldsymbol{K} + \sigma_0^2 \boldsymbol{I}\right)^{-1} \boldsymbol{k}(\boldsymbol{x}^*), & (7) \end{aligned}$$

where the elements of the vector $\boldsymbol{k}(\boldsymbol{x}^*)$ are defined as $k_i(\boldsymbol{x}^*) = \kappa(\boldsymbol{x}^*, \boldsymbol{x}_i)$, and the elements of matrix $\boldsymbol{K}$ are given by $K_{ij} = \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)$. The selection of hyper-parameters including the intrinsic noise σ_0^2 and kernel function parameters is usually achieved by optimizing the log marginal likelihood:

$$\begin{aligned} \log p(\boldsymbol{y} \mid \mathcal{D}) = &- \frac{1}{2} \boldsymbol{y}^{\mathrm{T}} \left(\boldsymbol{K} + \sigma_0^2 \boldsymbol{I}\right)^{-1} \boldsymbol{y} - \frac{1}{2} \log \left|\boldsymbol{K} + \sigma_0^2 \boldsymbol{I}\right| \\ &- \frac{n}{2} \log 2\pi. \end{aligned} \tag{8}$$

2.2.2 Acquisition Function. A good acquisition function balances between finding the worst performance (exploitation) and exploring in highly uncertain regions of the parameter space (exploration). Popular acquisition functions include: probability of improvement (PI), expected improvement (EI) and lower confidence bound (LCB) [14]. We adopt the approach of (9) which employs multiple acquisition functions with different levels of balancing between exploitation and exploration for added robustness:

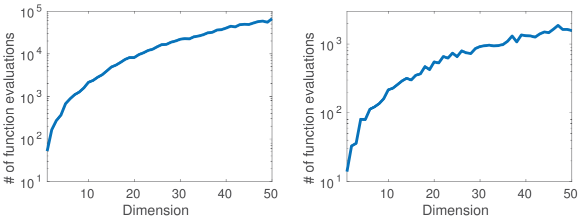
$$\begin{aligned} \alpha(\boldsymbol{x}; \mathcal{D}) &= \alpha_{pBO}(\boldsymbol{x}; \mathcal{D}, w) \\ &= (1 - w)\, \mu(\boldsymbol{x}; \mathcal{D}) - w \sigma(\boldsymbol{x}; \mathcal{D}), \end{aligned} \tag{9}$$

where w is the weighting parameter to balance exploitation and exploration for the acquisition function search direction.

3 CHALLENGES OF HIGH-DIMENSIONAL BO

A global optimization method assisted with a local gradient-free optimizer is usually to optimize the D-dimensional acquisition function. Such methods often suffer severely from the curse of dimensionality. We tested the optimization efficiency of DIRECT_L [3] and COBYLA [12] from the NLopt library [6] on a simple objective function:

$$y_{syn}(\boldsymbol{x}) = \frac{\|\boldsymbol{x} - \boldsymbol{c}\|_2}{\|\boldsymbol{c}\|_2}, \tag{10}$$



(a) DIRECT_L (b) COBYLA

Figure 2: Number of function evaluations per optimization for two optimization methods.

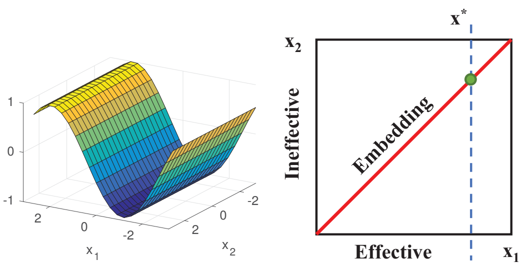
where c is a D-dimensional vector. Fig. 2 shows that the required number of function evaluations for both methods is super-linear in D. This suggests that in general when BO is applied to a black-box function the number of acquisition function evaluations can be much larger than D. The time complexity for evaluating simple acquisition functions like PI, EI and LCB is $O\left(N^2+ND\right)$, where N is the number of training examples. Therefore, the time complexity for optimizing the acquisition once is greater than $O\left(N^2D+ND^2\right)$ which is quadratic in D at minimum. Optimizing general non-convex acquisition functions in high-dimensions can be challenging. To force the completion, the number of acquisition function evaluations is upper bounded, leading to poor optimization quality. In addition, hyper-parameter tuning for GP models also suffers from high dimensionality.

4 PROPOSED HIGH-DIMENSIONAL BAYESIAN OPTIMIZATION

Our experimental studies have shown that the degradation of optimization solution quality and high time complexity of high-dimensional AMS circuits can make BO fail to detect rare design failures. We address this challenge by exploring random embedding to effectively reduce the dimensionality motivated by the fact typically only a subset of circuit parameters and parameter combinations have a significant impact on a target design performance.

4.1 Dimension Reduction: Random Embedding

Consider that the original D-dimensional parameter space has a d_e-dimensional effective linear subspace $\mathcal{V}$ such that for all $\boldsymbol{x}_e \in \mathcal{V}$ and $\boldsymbol{x}_u \in \mathcal{V}^{\perp}$, we have $y\left(\boldsymbol{x}_e+\boldsymbol{x}_u\right)=y\left(\boldsymbol{x}_e\right)$, where d_e is the minimum integer number satisfying this property. Intuitively, parametric variations in the subspace orthogonal to $\mathcal{V}$ with the lowest possible dimensionality d_e does not alter the performance value.



(a) 2D function surface (b) 1D embedding

Figure 3: Random embedding illustration.

As proven in [21], with a random matrix $\boldsymbol{A} \in \mathbb{R}^{D\times d}$ with entries independently sampled according to $\mathcal{N}(0,1)$, where embedding dimension $d \geq d_e$, for $\forall \boldsymbol{x} \in \mathbb{R}^D$, there exists a $\boldsymbol{z} \in \mathbb{R}^d$ such that $y(\boldsymbol{x}) = y(\boldsymbol{A}\boldsymbol{z})$ with probability 1. Therefore, the original high dimensional space can be embedded into a low dimensional space via a random matrix, resulting in a low dimension search space in Bayesian optimization. The optimum solution $\boldsymbol{x}^* \in \mathbb{R}^D$ can be found at some point $\boldsymbol{z}^* \in \mathbb{R}^d$, where $\boldsymbol{x}^* = \boldsymbol{A}\boldsymbol{z}^*$.

For example, the 2D objective function in Fig. 3 only depends on x_1. The 2D parameter space can be embedded into a 1D space (red solid line) along which the optimum solution can be found.

4.2 Proposed BO with Random Embedding

We define the failure search region for z as $\mathcal{Z} \subseteq \mathbb{R}^d$. Typically, the normalized failure search space Ω for $\boldsymbol{x}$ can be set as a bounded hyper-cube $[-1,1]^D$. The exact mapping of Ω in the embedding subspace may be complex, but can be well approximated by another bounded hyper-cube $[-\sqrt{d},\sqrt{d}]^d$ [21]. Now BO can operate in the low d-dimensional space defined by random embedding: both GP modeling and optimization of the acquisition function take place in terms of z. The sampling of training data for the GP model is confined in $\mathcal{Z}$. Each sampled $\boldsymbol{z} \in \mathcal{Z}$ is mapped to a $\boldsymbol{x} \in \Omega$ via the random matrix $\boldsymbol{A}$ by:

$$\boldsymbol{x} = p_{\Omega}\left(\boldsymbol{A}\boldsymbol{z}\right). \tag{11}$$

In case that $\boldsymbol{A}\boldsymbol{z}$ locates outside Ω, the projection operation $p_{\Omega}\left(\cdot\right)$ is performed to constrain the mapped $\boldsymbol{x}$ within Ω. Then the circuit performance at $\boldsymbol{x}$ is obtained using circuit simulation.

Algorithm 1: Proposed Bayesian optimization for failure detection in high dimension space

```
Input :Original function dimensionality $D$;
       Initial sample dataset $\mathcal{D}_0$;
       Simulation budget $n$; Batch size $n_b$;
       Preset $n_b$ weighting parameters $w_1, \ldots, w_{n_b}$;
       Objective function $y(\boldsymbol{x})$; Target specification $T$.
Output:Detected failure set $\mathcal{F}$.
Select an embedding dimension d from $\mathcal{D}_0$;
Sample a random matrix $\boldsymbol{A} \in \mathbb{R}^{D\times d}$;
Build the initial statistical model $p\left(y^* | \boldsymbol{z}^*,\ \mathcal{D}_0\right)$;
$\mathcal{F} \leftarrow \{\}$ ;
for $b \leftarrow 1$ to $n/n_b$ do
    for $i \leftarrow 1$ to $n_b$ do
        $z_{b,i} \leftarrow \arg\min_{z\in\mathcal{Z}} \alpha_{pBO}\left(z; \mathcal{D}_{b-1}, w_i\right)$;
        $y_{b,i} \leftarrow y\left(p_{\Omega}\left(\boldsymbol{A}\boldsymbol{z}_{b,i}\right)\right)$;
        if $y_{b,i} < T$ then
            $\mathcal{F} \leftarrow \left\{\mathcal{F}, \left(p_{\Omega}\left(\boldsymbol{A}\boldsymbol{z}_{b,i}\right), y_{b,i}\right)\right\}$;
        end
    end
    $\mathcal{D}_b \leftarrow \left\{\mathcal{D}_{b-1}, \left(z_{b,1}, y_{b,1}\right), \ldots, \left(z_{b,n_b}, y_{b,n_b}\right)\right\}$;
    Update statistical model $p\left(y^* | \boldsymbol{z}^*,\ \mathcal{D}_b\right)$;
end
return $\mathcal{F}$.
```

We summarize our Bayesian optimization algorithm using both random embedding technique and parallelizable acquisition function (9) as shown in Algorithm 1. With the proposed algorithm, the acquisition function optimization is executed in a low dimension space $\mathcal{Z} \subseteq \mathbb{R}^d$, which can be expected to have better optimization quality and efficiency. In addition, the GP model is trained under the low-dimensional space as well, resulting in more efficient GP training and evaluation.

4.3 Embedding Dimensionality Selection

While [21] provides the general theoretical principle of random embedding, it does not offer guidance for finding the effective dimensionality d_e. Selection of the embedding dimension d must balance two conflicting needs. An overly small d can lead to over-compression of the original parameters $\boldsymbol{x}$ and hence poor accuracy of the surrogate GP model, jeopardizing the robustness of failure detection. On the other hand, if d is too large, we can barely benefit from the dimension reduction brought by random embedding.

We propose the following data-efficient approach to select the embedding dimensionality prior to the BO based failure detection. For this, we collect a small training dataset to train multiple GP models with varying embedding dimensionalities. To share the same training dataset for all such GP models, the sampling of the training dataset takes place in the original D-dimensional parameter space, and the labels (circuit performance values) are queried using circuit simulation. Then, each sampled vector $\boldsymbol{x} \in \mathbb{R}^D$ is mapped to the corresponding vector $\boldsymbol{z} \in \mathbb{R}^d$ with embedding dimension d via pseudo inverse of the random embedding:

$$\boldsymbol{z} = \boldsymbol{A}^{\dagger}\boldsymbol{x} = \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)^{-1} \boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}. \tag{12}$$

The above procedure maps one common training dataset in $\boldsymbol{x}$ to a training set for each embedding dimension d such that a GP model with dimension d can be trained using the mapped data. We then use the mean-square error (MSE) to evaluate each GP model. If the dimensionality d is smaller than the unknown effective dimension d_e, we expect the MSE of the corresponding GP model would be large. We track the the variation of MSE as d increases. If the MSE stops decreasing at some dimension $\tilde{d}$, $\tilde{d}$ is likely to be just somewhat greater than d_e, and hence a good choice as the embedding dimension used for the sequential BO process. Since we only want to use a small amount of data to determine $\tilde{d}$, multiple, say T, GP models with different random matrices are trained for each d and their MSEs are averaged to minimize the variance of random embedding with small data. Embedding dimension selection is summarized in Algorithm 2.

Algorithm 2: Proposed embedding dimension selection.

```
Input  : Initial sample dataset $\mathcal{D}_0 = \{\boldsymbol{X}, \boldsymbol{y}\}$;
         Original function dimensionality $D$;
         Random matrix maximum trial count $T$.
Output : Embedded dimension $\tilde{d}$.
for $d \leftarrow 1$ to $D$ do
    for $i \leftarrow 1$ to $T$ do
        Sample a random matrix $\boldsymbol{A} \in \mathbb{R}^{D \times d}$;
        $\boldsymbol{A}^{\dagger} \leftarrow \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)^{-1} \boldsymbol{A}^{\mathrm{T}}$;
        Build statistical model $p\left(y^* \,|\boldsymbol{z}^*,\ \left\{\boldsymbol{A}^{\dagger}\boldsymbol{X}, \boldsymbol{y}\right\}\right)$;
        Compute $mse_i$ of the model given $\left\{\boldsymbol{A}^{\dagger}\boldsymbol{X}, \boldsymbol{y}\right\}$;
    end
    $MSE_d \leftarrow \frac{1}{T} \sum_{i=1}^{T} mse_i$;
end
Pick the smallest $\tilde{d}$ where MSE stops decreasing from the plot
  using $\{MSE_1 \cdots MSE_D\}$;
return $\tilde{d}$.
```

5 EXPERIMENTAL RESULTS

5.1 Experimental Setups

We demonstrated the effectiveness of the proposed Bayesian optimization approach with two circuits: an under-voltage lockout circuit [4] (19 dimensions) and a low-dropout regulator [8] (60 dimensions), as shown in Fig. 4 and 5, respectively. Both circuits are designed using a commercial 90nm CMOS technology design kit, and simulated remotely on the server with the 2.80GHz Intel(R) Xeon(R) E5-2680 v2 CPU.

The performances of two categories of techniques, i.e. sampling methods and Bayesian optimization, are studied. Under the first category, we employ Monte Carlo (MC) method and Scaled-Sigma Sampling (SSS) algorithm [18, 19], a state-of-the-art statistical sampling technique. The parameter variations of interest are bounded inside a large hyper-cube, which encloses a wide $\pm 4\sigma$ range for each parameter. To maximize the possibility of hitting rare failures within the large hyper-cube, uniform sampling distribution is adopted for MC. In the second category, Bayesian optimization approaches using different acquisition functions EI, PI, LCB [14] and the parallelizable multi-acquisition functions (pBO) [5] are selected to compare the proposed BO approach with random embedding. The BO methods were implemented in C++ under the BayesOpt [11] framework using DIRECT_L [3] for global optimization and COBYLA [12] for local optimization in NLopt library [6]. All the experiments were conducted on a workstation with a 3.50GHz Intel(R) Xeon(R) E5-1620 v4 CPU.

5.1.1 CMOS Under-voltage Lockout Circuit. The offset of the turn-off threshold voltage $|\Delta V_{THL}|$ is chosen as the interested verification target for the UVLO circuit, which may undergo dramatic

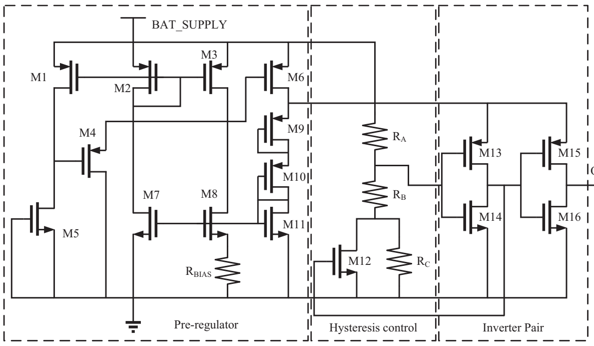


Figure 4: A CMOS under-voltage lockout circuit.

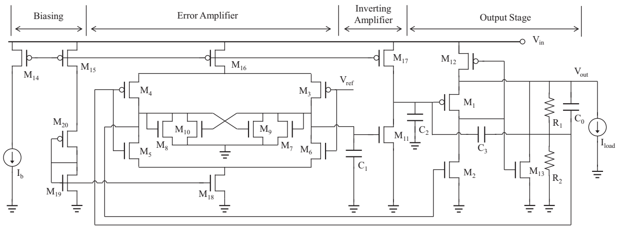


Figure 5: A low-dropout regulator.

fluctuations even with small parametric variations. The variation parameters considered are values of the three resistors and the channel lengths of all 16 transistors, resulting in a 19-dimensional parameter space. 5 initial samples are collected as the starting points for all the Bayesian optimization experiments, and another 95 samples are subsequently collected for failure detection.

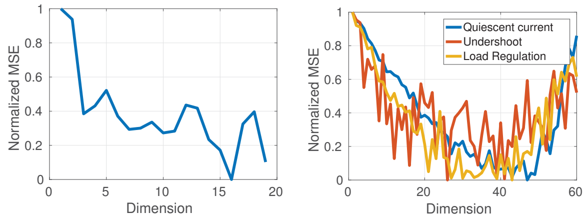
5.1.2 Low-dropout Regulator. Three specifications, quiescent current, undershoot and load regulation, are set as the verification targets for the LDO regulator. Three types of transistor-level variations are considered for all 20 transistors: channel length, threshold voltage and gate oxide thickness, resulting in a 60-dimensional verification problem. All experiments related to Bayesian optimization use the same 50 samples for the first GP model training and a simulation budget of 350 examples for the sequential experiment design later on.

5.2 Random Embedding Dimension Selection

To pick the embedding dimension $\tilde{d}$, Algorithm 2 is performed for both circuits. For this, 5 initial examples are used for the UVLO circuit, and 50 for the LDO. The GP model accuracy corresponding to various dimensions is presented in Fig.6, where the MSE results are normalized into the range of $[0, 1]$ for demonstration convenience. For the UVLO circuit, the minimum MSE is achieved at dimension 16, which however does not bring in much benefit from dimension reduction. Instead, we pick $\tilde{d}_{UVLO} = 8$, a good tradeoff between model accuracy and dimension reduction. For all the three specifications of the LDO, the MSE reaches the minimal level around dimension 30, therefore we set $\tilde{d}_{LDO} = 30$.

5.3 Failure Detection Effectiveness and Efficiency

As shown in Tables 1 and 2, the MC and SSS methods collect thousands to hundreds of thousands of simulation examples without detecting a single failure. This also indicates that the failures in these two circuits are extremely rare. Meanwhile, traditional acquisition functions like EI, PI and LCB or parallelizable Bayesian optimization (pBO) method cannot detect a single failure as well due to the inherent difficulty in applying BO in high-dimensional spaces. The proposed failure detection approach is the only method detecting failures for all specifications. The worst-case performance levels found by our method are much worse than the given target, while the statistical sampling methods and the more conventional Bayesian optimization methods are overly optimistic.



(a) UVLO (b) LDO

Figure 6: Random embedding dimension selection results.

Moreover, the number of simulation runs required by the proposed methods is much less than others. As presented in Tables 1 and 2, we only need 26 simulation data points to discover the first failure inside the 19-dimensional space for the UVLO circuit and hundreds of samples to detect the first failures in the 60-dimensional space for the LDO. The large reduction of simulation data brought by the proposed technique can be even more significant for rare failure detection of larger and more complex AMS circuits for which transistor-level simulation can be prohibitively expensive.

The runtime reported in Table 1 and 2 is the total runtime for Algorithm 1 including circuit simulation in a single thread configuration, i.e., no parallel mechanism is activated, which offers a clearer view of the runtime reduction provided by random embedding technique. As described earlier, since the original high-dimensional parameter space is embedded into a space of a lower dimensionality, the Gaussian process model can be trained at a much reduced cost, which speeds up both its posterior distribution evaluation and hyper-parameter tuning. Meanwhile, since the optimization of the acquisition function is also executed in the lower-dimensional space, the quality of optimization is improved and the number of function evaluations is greatly reduced, resulting in significantly less runtime compared to other Bayesian optimization methods. The runtime for Algorithm 2 is typically less than one minute with small sample size, which can be ignored compared to expensive simulation cost.

6 CONCLUSION

In this paper, we present a high-dimensional Bayesian optimization procedure for rare failure detection of analog/mixed-signal circuits. We utilized random embedding techniques to remove redundant features and reduce the dimensionality of Bayesian optimization search space, resulting highly-efficient failure detection. Our experimental results demonstrate that the proposed Bayesian optimization is capable of detecting rare failures in high dimension variation space with only hundreds of simulation samples, while both statistical sampling and Bayesian optimization techniques with traditional acquisition functions completely miss.

ACKNOWLEDGMENTS

This material is based upon work supported by the Semiconductor Research Corporation (SRC) through Texas Analog Center of Excellence at the University of Texas at Dallas (Task ID:2712.004 and 2810.024).

REFERENCES

[1] Eric Brochu, Vlad M. Cora, and Nando De Freitas. 2010. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv preprint arXiv:1012.2599* (Dec 2010).

[2] Henry Chang and Ken Kundert. 2007. Verification of Complex Analog and RF IC Designs. *Proc. IEEE* 95, 3 (March 2007), 622–639.

Table 1: Failure detection result comparison for the UVLO circuit verification (19 dimension).

Spec	Target	Method	# Sim	Worst Case	1st Failure Hit	Runtime
$\|\Delta V_{THL}\|$	0.9V	MC	20,000	0.86V	-	4h22m07s
		SSS	1,000	0.15V	-	13m24s
		EI	$5_{init} + 95_{seq}$	0.16V	-	8m30s
		PI	$5_{init} + 95_{seq}$	0.04V	-	7m56s
		LCB	$5_{init} + 95_{seq}$	0.17V	-	7m40s
		pBO	$5_{init} + 5 \times 19_{batch}$	0.14V	-	9m01s
		This work	$5_{init} + 5 \times 19_{batch}$	**0.95V**	**26**	**5m32s**

Table 2: Failure detection result comparison for the LDO regulator verification (60 dimension).

Spec	Target	Method	# Sim	Worst Case	1st Failure Hit	Runtime
Quiescent current	12mA	MC	649,000	11.6mA	-	160h25m12s
		SSS	6,000	8.2mA	-	1h38m41s
		EI	$50_{init} + 350_{seq}$	7.0mA	-	6h46m13s
		PI	$50_{init} + 350_{seq}$	7.2mA	-	6h00m23s
		LCB	$50_{init} + 350_{seq}$	8.0mA	-	6h33m42s
		pBO	$50_{init} + 5 \times 70_{batch}$	7.0mA	-	8h03m19s
		This work	$50_{init} + 5 \times 70_{batch}$	**12.7mA**	**231**	**1h57m05s**
Undershoot	0.40V	MC	649,000	0.39V	-	160h25m12s
		SSS	6,000	0.19V	-	1h38m41s
		EI	$50_{init} + 350_{seq}$	0.20V	-	6h45m11s
		PI	$50_{init} + 350_{seq}$	0.20V	-	6h51m05s
		LCB	$50_{init} + 350_{seq}$	0.18V	-	6h23m55s
		pBO	$50_{init} + 5 \times 70_{batch}$	0.14V	-	7h51m14s
		This work	$50_{init} + 5 \times 70_{batch}$	**0.51V**	**87**	**2h01m52s**
Load regulation	50.0%	MC	649,000	47.2%	-	160h25m12s
		SSS	6,000	22.6%	-	1h38m41s
		EI	$50_{init} + 350_{seq}$	25.1%	-	6h49m51s
		PI	$50_{init} + 350_{seq}$	9.2%	-	6h35m03s
		LCB	$50_{init} + 350_{seq}$	28.4%	-	6h38m14s
		pBO	$50_{init} + 5 \times 70_{batch}$	17.3%	-	7h50m02s
		This work	$50_{init} + 5 \times 70_{batch}$	**55.0%**	**302**	**1h55m14s**

[3] Jorg M. Gablonsky and Carl T. Kelley. 2001. A Locally-Biased form of the DIRECT Algorithm. *Journal of Global Optimization* 21, 1 (Sep 2001), 27–37.

[4] Mohammad R. Hoque and Simon S. Ang. 2008. A CMOS under-voltage lockout circuit. In *World Congress on Engineering and Computer Science (WCECS)*. 1–4.

[5] Hanbin Hu, Peng Li, and Jianhua Z. Huang. 2018. Parallelizable Bayesian Optimization for Analog and Mixed-signal Rare Failure Detection with High Coverage. In *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*. Article 98, 8 pages.

[6] Steven G. Johnson. 2014. The NLopt nonlinear-optimization package [Software]. (2014). http://ab-initio.mit.edu/nlopt

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

[8] Suming Lai and Peng Li. 2012. A fully on-chip area-efficient CMOS low-dropout regulator with fast load regulation. *Analog Integrated Circuits and Signal Processing* 72, 2 (August 2012), 433–450.

[9] Honghuang Lin, Asad M. Khan, and Peng Li. 2017. Statistical circuit performance dependency analysis via sparse relevance kernel machine. In *IEEE International Conference on IC Design and Technology (ICICDT)*. 1–4.

[10] Honghuang Lin and Peng Li. 2016. Relevance vector and feature machine for statistical analog circuit characterization and built-in self-test optimization. In *ACM/IEEE Design Automation Conference (DAC)*. 1–6.

[11] Ruben Martinez-Cantin. 2014. BayesOpt: A Bayesian Optimization Library for Nonlinear Optimization, Experimental Design and Bandits. *Journal of Machine Learning Research* 15 (2014), 3915–3919.

[12] M. J. D. Powell. 1998. Direct search algorithms for optimization calculations. *Acta Numerica* 7 (1998), 287–336.

[13] Carl E. Rasmussen and Christopher K. I. Williams. 2006. *Gaussian process for machine learning.* MIT Press.

[14] Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando de Freitas. 2016. Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proc. IEEE* 104, 1 (Jan 2016), 148–175.

[15] Amith Singhee and Rob A. Rutenbar. 2009. Statistical Blockade: Very Fast Statistical Simulation and Modeling of Rare Circuit Events and Its Application to Memory Design. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 28, 8 (Aug 2009), 1176–1189.

[16] Jasper Snoek, Hugo Larochelle, and Ryan P. Adams. 2012. Practical Bayesian optimization of machine learning algorithms. In *International Conference on Neural Information Processing Systems (NIPS)*. 2951–2959.

[17] Niranjan Srinivas, Andreas Krause, Sham Kakade, and Matthias Seeger. 2010. Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design. In *Proceedings of the International Conference on International Conference on Machine Learning (ICML)*. 1015–1022.

[18] Shupeng Sun, Xin Li, Hongzhou Liu, Kangsheng Luo, and Ben Gu. 2013. Fast statistical analysis of rare circuit failure events via scaled-sigma sampling for high-dimensional variation space. In *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*. 478–485.

[19] Shupeng Sun, Xin Li, Hongzhou Liu, Kangsheng Luo, and Ben Gu. 2015. 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 (July 2015), 1096–1109.

[20] Kevin Swersky, Jasper Snoek, and Ryan P. Adams. 2013. Multi-Task Bayesian Optimization. In *Advances in Neural Information Processing Systems 26*. Curran Associates, Inc., 2004–2012.

[21] Ziyu Wang, Masrour Zoghi, Frank Hutter, David Matheson, and Nando De Freitas. 2013. Bayesian Optimization in High Dimensions via Random Embeddings. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*. 1778–1784.