Fast Statistical Analysis of Rare Circuit Failure Events via Bayesian Scaled-Sigma Sampling for High-Dimensional Variation Space

Shupeng Sun and Xin Li

Electrical & Computer Engineering Department, Carnegie Mellon University, Pittsburgh, PA, USA, 15213

***Abstract*—Accurately estimating the rare failure events of nanoscale ICs in a high-dimensional variation space is extremely challenging. In this paper, we propose a novel Bayesian scaled-sigma sampling (BSSS) technique to address this technical challenge. BSSS can be considered as an extension of the traditional scaled-sigma sampling (SSS) approach. The key idea is to explore the "similarity" between different SSS models fitted at different design stages and encode it as our prior knowledge. Bayesian model fusion is then adopted to fit the SSS model with consideration of the prior knowledge. A sense amplifier example designed in a 45 nm CMOS process is used to demonstrate the efficacy of BSSS. Experimental results demonstrate that BSSS achieves superior accuracy over the conventional SSS and minimum-norm importance sampling approaches when a few hundred random variables are used to model process variations.**

I. Introduction

The advancement of the nanoscale integrated circuit (IC) technology brings about large-scale process variations, which makes robust IC design a challenging task [1]. An advanced microprocessor chip typically integrates millions or even billions of circuit components (e.g., SRAM bit cells), and it must function correctly in the presence of large-scale process variations. In order to achieve sufficiently high yield, the failure rate of each circuit component must be extremely small. For example, the failure rate of an SRAM bit cell must be smaller than 10^{-8} since we have millions of SRAM bit cells on a typical microprocessor chip. The failure rate is so small that efficiently and accurately analyzing the rare circuit failure events becomes an important-yet-challenging task for the IC design community.

To solve this problem, numerous approaches and techniques have been proposed for failure rate estimation of SRAM bit cells where only a small number of (e.g., 10~20) independent random variables are used to model process variations and, hence, the dimensionality of the corresponding variation space is low [2]-[5]. It has been demonstrated in the literature [6]-[7] that rare failure rate estimation in a high-dimensional variation space (e.g., hundreds of independent random variables to model the device-level variations for a sense amplifier) becomes more and more important. Unfortunately, these traditional techniques [2]-[5] cannot be efficiently applied when the dimensionality of the variation space becomes high. Realizing this limitation, two approaches [6]-[7] have recently been proposed to address the technical challenge posed by high dimensionality.

Though only a few thousand simulations are required by the state-of-the-art techniques for rare failure probability estimation [6]-[7], tens of thousands of simulations in total may be needed when designing a circuit component. To understand the reason, let us assume that we are designing a sense amplifier (SA), and applying the traditional approach [6], referred to as scaled-sigma sampling (SSS), to estimate the failure probability. If the estimated failure probability meets our design specification, the design process is complete. Otherwise, we need to improve our SA design, and re-run SSS. Generally speaking, we need a few iterations before we converge to the final design. Since SSS spends a few thousand simulations for each design candidate, tens of thousands of simulations may be required over the entire design process, which can be extremely expensive.

To further reduce the simulation cost, we propose a novel *Bayesian scaled-sigma sampling* (BSSS) approach which can be considered as an extension of the traditional SSS approach. SSS is a model-based approach. By studying SSS, we observe that a number of coefficients of the SSS model for an early design candidate are similar to those for a late design candidate. Motivated by this observation, we propose to explore the "similarity" between different SSS models fitted for different design candidates, and encode such "similarity" as our *prior* knowledge. Next, we apply Bayesian model fusion (BMF) [8] to fit the SSS model with the *prior*. Our numerical experimental results demonstrate that BSSS can achieve superior estimation accuracy over other traditional approaches [5]-[6] when a few hundred independent random variables are used to model process variations.

The remainder of this paper is organized as follows. In Section II, we briefly review the background of the traditional SSS approach, and then propose the BSSS method in Section III. An SA example is presented to demonstrate the efficacy of BSSS in Section IV. Finally, we conclude in Section V.

II. Background

Without loss of generality, we use a vector

$$\mathbf{x} = [x_1 \quad x_2 \quad \cdots \quad x_M]^T \tag{1}$$

to denote an M-dimensional random variable modeling process variations and $f(\mathbf{x})$ to denote the joint probability density function (PDF) of $\mathbf{x}$. The PDF $f(\mathbf{x})$ is typically modeled as a multivariate Gaussian distribution with independent, zero-mean and unit-variance variables [6]

$$f(\mathbf{x}) = \prod_{m=1}^{M} \left[\exp\left(-x_m^2/2\right) \Big/ \sqrt{2\pi} \right]. \tag{2}$$

Any correlated random variables that are jointly Gaussian can

978-1-4799-8682-8/15/$31.00 ©2015 IEEE

be transformed to the independent random variables by principal component analysis (PCA) [9].

The failure rate P_f of a circuit can be represented as

$$P_f = \int_{\Omega} f(\mathbf{x}) \cdot d\mathbf{x} \tag{3}$$

where Ω denotes the failure region in the variation space. For our application of rare failure event analysis, the conventional brute-force Monte Carlo (MC) analysis that directly samples $f(\mathbf{x})$ is extremely expensive, as explained in [6]-[7].

To address the efficiency issue, SSS was proposed in [6]. It samples a distorted PDF

$$g(\mathbf{x}) = \prod_{m=1}^{M} \left[\exp\left(-x_m^2 / 2s^2\right) \Big/ \left(\sqrt{2\pi} \cdot s\right) \right] \tag{4}$$

where s denotes the standard deviation of $g(\mathbf{x})$. SSS conceptually increases the magnitude of process variations and, hence, is more likely to obtain a random sample that can reach the rare failure region Ω. The "scaled" failure rate P_g associated with the "scaled" PDF $g(\mathbf{x})$ can be written as

$$P_g = \int_{\Omega} g(\mathbf{x}) \cdot d\mathbf{x} \,. \tag{5}$$

By scaling up the standard deviation, P_g becomes significantly larger than P_f. Therefore, SSS can efficiently estimate P_g with a small number of samples as long as the scaling factor s is sufficiently large.

To estimate the rare failure rate P_f defined in (3), SSS applies an analytical model to map P_g to P_f. It partitions the failure region Ω into a large number of identical hyper-rectangles and approximates (5) by a summation

$$P_g \approx \sum_{k \in \Omega} g\left[\mathbf{x}^{(k)}\right] \cdot \Delta\mathbf{x} \tag{6}$$

where $\Delta\mathbf{x}$ denotes the volume of a hyper-rectangle, and $\{k;\ k \in \Omega\}$ represents the set of all hyper-rectangles that are inside Ω. Substitute (4) into (6) and take the logarithm on both sides of (6), yielding

$$\log P_g \approx \alpha + \beta \cdot \log s + \gamma / s^2 \tag{7}$$

where

$$\alpha = \log\left[\Delta\mathbf{x} \Big/ \left(\sqrt{2\pi}\right)^M \right] \tag{8}$$

$$\beta = -M \tag{9}$$

$$\gamma = \max_{k \in \Omega} \left[-\left\| \mathbf{x}^{(k)} \right\|_2^2 / 2 \right]. \tag{10}$$

More details about the aforementioned model template can be found in [6].

Next, SSS chooses a number of scaling factors $\{s_q;\ q = 1, 2, \cdots, Q\}$ and estimates their corresponding scaled failure rates $\{P_{g,q};\ q = 1, 2, \cdots, Q\}$. Once $\{(s_q, P_{g,q});\ q = 1, 2, \cdots, Q\}$ are available, the three model coefficients (i.e., α, β, and γ) are fitted by maximum likelihood estimation (MLE), and the rare failure rate P_f is calculated by setting s to 1 in (7)

$$P_f^{SSS} = \exp(\alpha + \gamma). \tag{11}$$

III. Bayesian Scaled-Sigma Sampling

Studying the theoretical definition of the three model coefficients shown in (8)-(10), we have two interesting observations:

- **O1**: The first two model coefficients (i.e., α and β) strongly depend on the dimensionality (i.e., M) of the variation space, but are independent of the location of the failure region Ω.
- **O2**: The third model coefficient (i.e., γ) is determined by the hyper-rectangle that is inside the failure region Ω and is closest to the origin $\mathbf{0}$. Alternatively speaking, γ strongly depends on the location of the failure region Ω.

To fully understand the implication of these observations, let us consider two design candidates (i.e., an early design and a late design). We assume that the early design does not meet the performance specification and, hence, the late design is created from the early design by tuning its design variables (e.g., transistor sizes) to improve performance. The SSS models for these two design candidates are as follows

$$\log P_{E,g} \approx \alpha_E + \beta_E \cdot \log s + \gamma_E / s^2 \tag{12}$$

$$\log P_{L,g} \approx \alpha_L + \beta_L \cdot \log s + \gamma_L / s^2 \tag{13}$$

where $[\alpha_E\ \beta_E\ \gamma_E]$ and $[\alpha_L\ \beta_L\ \gamma_L]$ denote the model coefficients of the early and late designs respectively. Note that $\boldsymbol{\theta}_E = [\alpha_E\ \beta_E\ \gamma_E]$ is already fitted for the early design before we start to work on the late design. Our objective is to efficiently fit $\boldsymbol{\theta}_L = [\alpha_L\ \beta_L\ \gamma_L]$ in order to estimate the rare failure rate for the late design.

In this example, the dimensionalities of the early and late designs are likely to be similar or even identical. However, their failure regions can be different. Based on the two observations (i.e., O1 and O2), we can expect that

- α_E (or β_E) and α_L (or β_L) are likely to be similar, and
- γ_E and γ_L can be substantially different.

Such a similarity between the model coefficients has not been explored by the conventional SSS approach. In this paper, we propose a novel Bayesian scaled-sigma sampling (BSSS) approach to take advantage of the aforementioned knowledge. In particular, BSSS encodes the "similarity" between $\boldsymbol{\theta}_E$ and $\boldsymbol{\theta}_L$ as a *prior* distribution $pdf(\boldsymbol{\theta}_L)$, and applies BMF [8] to solve $\boldsymbol{\theta}_L$ by maximum-a-posteriori (MAP) estimation

$$\max_{\boldsymbol{\theta}_L} \quad pdf\left(\boldsymbol{\theta}_L \middle| \mathbf{D}\right) \propto pdf(\boldsymbol{\theta}_L) \cdot pdf\left(\mathbf{D} \middle| \boldsymbol{\theta}_L\right) \tag{14}$$

where $\mathbf{D}$ denotes the simulation data collected for the late design and $pdf(\mathbf{D} \mid \boldsymbol{\theta}_L)$ denotes the likelihood of observing the data $\mathbf{D}$.

The key difference between BSSS and SSS lies in the fact that BSSS maximizes the product of the prior distribution $pdf(\boldsymbol{\theta}_L)$ and the likelihood $pdf(\mathbf{D} \mid \boldsymbol{\theta}_L)$ by MAP, while SSS only maximizes the likelihood $pdf(\mathbf{D} \mid \boldsymbol{\theta}_L)$ by MLE. As long as the prior distribution $pdf(\boldsymbol{\theta}_L)$ is properly defined, MAP can reduce the amount of required simulation data and, hence, the model fitting cost without surrendering any accuracy, as demonstrated in [8]. In other words, BSSS can be more efficient than SSS if $pdf(\boldsymbol{\theta}_L)$ is appropriately defined.

Our proposed BSSS method consists of two major steps:

(i) constructing a prior distribution $pdf(\mathbf{\theta}_L)$ based on the similarity between $\mathbf{\theta}_E$ and $\mathbf{\theta}_L$, and (ii) optimally determining $\mathbf{\theta}_L$ by MAP estimation. In what follows, we will describe these two steps in detail.

A. Prior Knowledge Definition

Based on our knowledge of "similarity", we define α_L and β_L as two Gaussian random variables

$$\begin{aligned} pdf(\alpha_L) &= Gauss(\alpha_E, \sigma_\alpha^2) \\ pdf(\beta_L) &= Gauss(\beta_E, \sigma_\beta^2) \end{aligned} \tag{15}$$

where α_E and β_E are the means of α_L and β_L respectively, and σ_α and σ_β denote the standard deviations of α_L and β_L respectively. In (15), the standard deviations σ_α and σ_β can be optimally estimated by MLE, as will be further discussed at the end of this sub-section. Because a Gaussian distribution peaks at its mean value, the model coefficient α_L (or β_L) for the late design is unlikely to substantially deviate from its mean value α_E (or β_E) which is the model coefficient for the early design. Restating in words, the prior distributions defined in (15) attempt to capture the similarity between α_E (or β_E) and α_L (or β_L).

On the other hand, since γ_E and γ_L can be substantially different, we have no prior knowledge about γ_L. Hence, we can only define a non-informative prior [9] for γ_L

$$pdf(\gamma_L) = \begin{cases} 1/(u-l) & \gamma_L \in [l,u] \\ 0 & \text{else} \end{cases} \tag{16}$$

where l and u denote the lower and upper bounds for γ_L respectively. When defining $pdf(\gamma_L)$ in (16), we should choose l to be sufficiently small and u to be sufficiently large. As such, γ_L is uniformly distributed over a large range, implying that we do not know the value of γ_L in advance.

Next, we need to combine (15) and (16) to define the joint distribution for $\mathbf{\theta}_L = [\alpha_L\ \beta_L\ \gamma_L]$. Since we do not know the correlation among α_L, β_L, and γ_L, we simply assume that they are independent

$$pdf(\mathbf{\theta}_L) = pdf(\alpha_L) \cdot pdf(\beta_L) \cdot pdf(\gamma_L). \tag{17}$$

The correlation information will be learned from the data $\mathbf{D}$, when applying MAP estimation in the next sub-section.

Given the prior distribution in (17), we can now apply MLE to determine the values for σ_α and σ_β in (15)

$$\max_{\sigma_\alpha, \sigma_\beta} \quad pdf(\mathbf{D} | \sigma_\alpha, \sigma_\beta). \tag{18}$$

Eq. (18) can be re-written as

$$\max_{\sigma_\alpha, \sigma_\beta} \quad \int_{\mathbf{\theta}_L} pdf(\mathbf{D}|\mathbf{\theta}_L) \cdot pdf(\mathbf{\theta}_L | \sigma_\alpha, \sigma_\beta) \cdot d\mathbf{\theta}_L \tag{19}$$

where $pdf(\mathbf{D} \mid \mathbf{\theta}_L)$ is the likelihood function and $pdf(\mathbf{\theta}_L \mid \sigma_\alpha, \sigma_\beta)$ is the prior distribution in (17). Due to the page limit, more details about solving (19) are omitted here. Once σ_α and σ_β are found, the prior distribution in (17) is fully determined.

B. Maximum-A-Posteriori Estimation

Once the prior distribution in (17) is defined, we apply BMF [8] to solve $\mathbf{\theta}_L$ by MAP estimation, as shown in (14). Substituting (16)-(17) into (14) yields

$$\max_{\mathbf{\theta}_L} pdf(\mathbf{\theta}_L|\mathbf{D}) \propto \begin{cases} pdf(\alpha_L) \cdot pdf(\beta_L) \cdot pdf(\mathbf{D}|\mathbf{\theta}_L) & \gamma_L \in [l,u] \\ 0 & \text{else} \end{cases} \tag{20}$$

where $\mathbf{\theta}_L = [\alpha_L\ \beta_L\ \gamma_L]$. It can be shown that Eq. (20) is a convex optimization problem and its solution can be found by inspecting the first-order optimality condition [10].

Once the MAP solution $\mathbf{\theta}_L^{MAP} = [\alpha_L^{MAP}\ \beta_L^{MAP}\ \gamma_L^{MAP}]$ for the late design is obtained, the rare failure rate P_f is calculated by setting the scaling factor s to 1 in the model template (7)

$$P_f^{BSSS} = \exp\left(\alpha_L^{MAP} + \gamma_L^{MAP}\right). \tag{21}$$

C. Confidence Interval Estimation via Bootstrap

To quantitatively assess the accuracy of our proposed BSSS method, we apply bootstrap [11] to accurately estimate the confidence interval of our BSSS estimator in (21). The key idea of bootstrap is to re-generate a large number of datasets $\{\mathbf{D}^{(n)};\ n = 1, 2, \cdots, N^{BOOT}\}$ based on a statistical model without running additional transistor-level simulations. These datasets $\{\mathbf{D}^{(n)};\ n = 1, 2, \cdots, N^{BOOT}\}$ are then used to repeatedly run BSSS for N^{BOOT} times and we get N^{BOOT} different failure rates $\{P_f^{BSSS(n)};\ n = 1, 2, \cdots, N^{BOOT}\}$. Based on $\{P_f^{BSSS(n)};\ n = 1, 2, \cdots, N^{BOOT}\}$, the statistics (hence, the confidence interval) of the estimator P_f^{BSSS} can be accurately estimated. More details about the bootstrap approach can be found in [6].

D. Summary

Algorithm 1: **Bayesian Scaled-Sigma Sampling** (**BSSS**)

1. Start from the model coefficients $\mathbf{\theta}_E$ for the early design.
2. Collect the simulation data $\mathbf{D}$ for the late design.
3. Form the prior distribution for $\mathbf{\theta}_L$ by using (15)-(17).
4. Solve $\mathbf{\theta}_L$ by the MAP estimation in (20).
5. Calculate the failure rate P_f by the estimator P_f^{BSSS} in (21).
6. Generate N^{BOOT} datasets $\{\mathbf{D}^{(n)};\ n = 1, 2, \cdots, N^{BOOT}\}$ by re-sampling [6].
7. For each dataset $\mathbf{D}^{(n)}$ where $n \in \{1, 2, \cdots, N^{BOOT}\}$, repeat Step 3~5 to calculate the failure rate $P_f^{BSSS(n)}$.
8. Based on $\{P_f^{BSSS(n)};\ n = 1, 2, \cdots, N^{BOOT}\}$, estimate the confidence interval of the estimator P_f^{BSSS}.

Algorithm 1 summarizes the simplified flow of the proposed BSSS approach. It assumes that the model coefficients $\mathbf{\theta}_E$ for the early design are already given, and our objective is to estimate the rare failure rate for a late design. If the prior knowledge encoded by $\mathbf{\theta}_E$ is accurate, BSSS is expected to offer superior accuracy over the conventional SSS method, as will be demonstrated by the numerical example in the next section.

IV. Experimental Results

Shown in Fig. 1 is the simplified circuit block diagram for a sense amplifier (SA) designed in a 45 nm CMOS process. The SA consists of 45 transistors and each transistor consists of a number of multipliers. The random mismatch of each multiplier is modeled by 4 independent Gaussian random variables in the process design kit. In total, there are 536 independent Gaussian random variables. The BL and BLB

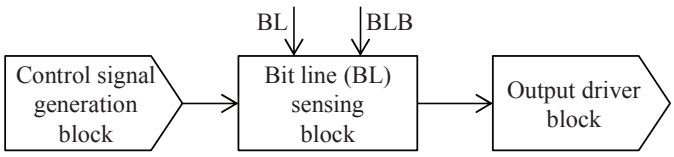

Fig. 1. Simplified circuit block diagram is shown for a sense amplifier.

voltages are initially set to 1.1V and 1.2V respectively. If the output of SA is 0, the SA is considered as "PASS". Otherwise, it is considered as "FAIL".

We start from an initial SA design, and run SSS with 10^4 transistor-level simulations to estimate its failure rate. The estimated failure rate by SSS is 7.3×10^{-5}, which is relatively high and does not meet our specification. Hence, we need to further improve the SA design to reduce its failure rate.

After tuning the transistor sizes, we obtain a second design. For testing and comparison purposes, we run four different approaches to estimate the failure rate of the second design:

- **MC**: the traditional brute-force Monte Carlo approach that is used to estimate the "golden" failure rate.
- **MNIS**: the traditional minimum-norm importance sampling approach [5] for which 2000 simulations are used to search the variation space and find the "distorted" distribution for sampling.
- **SSS** [6]: the traditional SSS approach where 5 different scaling factors are empirically chosen to fit the model coefficients α, β, and γ in (7).
- **BSSS**: the proposed BSSS approach as shown in Algorithm 1.

For the second SA design, we first run MC with 3.5×10^6 random samples. The estimated failure rate is 7.1×10^{-6}, which is considered as the "golden" failure rate of the second design. Next, we run MNIS, SSS, and BSSS with different numbers of simulations. The estimated failure rates P_f and the 95% confidence intervals $[P_f^L \; P_f^U]$ by these three approaches are shown in Table I.

Studying Table I, we have several observations. First, MNIS does not predict the failure rate or the 95% confidence interval accurately even with 10^4 simulations. We believe that MNIS cannot find a good "distorted" distribution in this high-dimensional example and, hence, misses the most important failure region. Second, both SSS and BSSS estimate the failure rate and the 95% confidence interval more accurately than MNIS.

Last, but more importantly, BSSS achieves significantly enhanced accuracy over SSS. From Table I, we can observe that the 95% confidence interval of BSSS with 6000 simulations is substantially narrower than that of SSS with 10^4 simulations. Alternatively speaking, our proposed BSSS approach achieves more than 1.7× runtime speedup over the state-of-the-art failure rate estimation technique (i.e., SSS) in this example.

Table I
FAILURE RATE P_f AND 95% CI $[P_f^L, P_f^U]$ ESTIMATED BY MNIS, SSS, AND BSSS ("GOLDEN" FAILURE RATE = 7.1×10^{-6})

# of Sims		6000	7000	8000	9000	10000
MNIS [5]	P_f^L	0	0	0	0	0
	P_f	1.5×10^{-15}	1.2×10^{-15}	1.1×10^{-13}	4.1×10^{-13}	3.6×10^{-13}
	P_f^U	3.1×10^{-15}	2.5×10^{-15}	3.1×10^{-13}	1.1×10^{-12}	9.2×10^{-13}
SSS [6]	P_f^L	9.0×10^{-7}	9.4×10^{-7}	6.2×10^{-7}	4.5×10^{-7}	5.1×10^{-7}
	P_f	4.4×10^{-5}	3.8×10^{-5}	2.0×10^{-5}	1.1×10^{-5}	1.0×10^{-5}
	P_f^U	2.3×10^{-4}	2.0×10^{-4}	1.7×10^{-4}	8.8×10^{-5}	7.9×10^{-5}
BSSS	P_f^L	2.4×10^{-6}	1.8×10^{-6}	1.7×10^{-6}	1.5×10^{-6}	1.5×10^{-6}
	P_f	1.3×10^{-5}	9.0×10^{-6}	8.0×10^{-6}	7.0×10^{-6}	6.2×10^{-6}
	P_f^U	4.8×10^{-5}	4.1×10^{-5}	2.8×10^{-5}	2.4×10^{-5}	2.0×10^{-5}

V. CONCLUSION

In this paper, we propose a novel BSSS approach to accurately estimate the rare failure rates for nanoscale ICs in a high-dimensional variation space. BSSS explores the "similarity" between different SSS models fitted at different design stages, and encodes it as our prior knowledge. Next, the SSS model is fitted by using MAP estimation with consideration of the prior knowledge. Experimental results demonstrate that the proposed BSSS approach achieves superior accuracy over the traditional MNIS and SSS approaches when the dimensionality of the variation space is more than a few hundred.

ACKNOWLEDGMENTS

This work has been supported by the National Science Foundation under contract CCF–1148778.

REFERENCES

[1] B. Calhoun, et al., "Digital circuit design challenges and opportunities in the era of nanoscale CMOS," *Proc. IEEE*, vol. 96, no. 2, pp. 343-365, Feb. 2008.

[2] R. Kanj, R. Joshi and S. Nassif, "Mixture importance sampling and its application to the analysis of SRAM designs in the presence of rare failure events," *IEEE DAC*, pp. 69-72, 2006.

[3] C. Gu and J. Roychowdhury, "An efficient, fully nonlinear, variability aware non-Monte-Carlo yield estimation procedure with applications to SRAM cells and ring oscillators," *IEEE ASP-DAC*, pp. 754-761, 2008.

[4] A. Singhee and R. Rutenbar, "Statistical blockade: very fast statistical simulation and modeling of rare circuit events, and its application to memory design," *IEEE TCAD*, vol. 28, no. 8, pp. 1176-1189, Aug. 2009.

[5] M. Qazi, at al., "Loop flattening and spherical sampling: highly efficient model reduction techniques for SRAM yield analysis," *IEEE DATE*, pp. 801-806, 2010.

[6] S. Sun, et al., "Fast statistical analysis of rare circuit failure events via scaled-sigma sampling for high-dimensional variation space," *IEEE ICCAD*, pp. 478-485, 2013.

[7] S. Sun and X. Li, "Fast statistical analysis of rare circuit failure events via subset simulation in high-dimensional variation space," *IEEE ICCAD*, pp. 324-331, 2014.

[8] F. Wang, et al., "Bayesian model fusion: large-scale performance modeling of analog and mixed-signal circuits by reusing early-stage data," *IEEE DAC*, 2013.

[9] C. Bishop, *Pattern Recognition and Machine Learning*. Prentice Hall, 2007.

[10] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2009.

[11] B. Efron and R. Tibshirnani, *An Introduction to the Bootstrap*. Chapman & Hall/CRC, 1993.