Efficient Multivariate Moment Estimation via Bayesian Model Fusion for Analog and Mixed-Signal Circuits

Qicheng Huang[1], Chenlei Fang[1], Fan Yang[1,*], Xuan Zeng[1,*] and Xin Li[1,2]
[1]State Key Lab of ASIC & System, Microelectronics Department, Fudan University, Shanghai, P. R. China
[2]Electrical & Computer Engineering Department, Carnegie Mellon University, Pittsburgh, PA, USA

ABSTRACT

A critical-yet-challenging problem of analog/mixed-signal circuit validation in either pre-silicon or post-silicon stage is to estimate the parametric yield of the performances. In this paper, we propose a novel Bayesian model fusion method for efficient multivariate moment estimation of multiple correlated performance metrics by borrowing the prior knowledge from the early stage. The key idea is to model the multiple performance metrics as a jointly Gaussian distribution and encode the prior knowledge as a normal-Wishart distribution according to the theory of conjugate prior. The late-stage multivariate moments can be accurately estimated by Bayesian inference with very few late-stage samples. Several circuit examples demonstrate that the proposed method can achieve up to 16× cost reduction over the traditional method without surrendering any accuracy.

1. INTRODUCTION

The continuous scaling of integrated circuits (ICs) leads to severe process variations. Process variations, in terms of doping profiles, interconnect widths, channel lengths, etc., lead to large-scale variations in the corresponding electrical parameters (e.g., resistance, capacitance, threshold voltage, etc.) of CMOS transistors and metal interconnects. These variations would consequently impact the parametric yield of analog and mixed-signal (AMS) circuits. Hence, accurate yield estimation is one of the critical tasks for both pre-silicon verification and post-silicon validation of AMS circuits in order to improve the circuit performance and/or reduce the manufacturing cost at advanced technology nodes [1-4].

Recently, post-silicon tuning and self-healing techniques have been proposed to address the yield loss posed by process variations. These methods adaptively adjust a number of tunable knobs (e.g., bias current) to meet the performance requirement after manufacturing [3-4]. They pose great challenges to yield estimation, as a large number of samples must be collected for a highly complex, tunable circuit by either circuit simulation (for pre-silicon verification) or silicon measurement (for post-silicon validation). Both circuit simulation and silicon measurement are time-consuming. For example, one single post-layout simulation of a large-scale AMS circuit such as PLL or SRAM could take several days to finish. On the other hand, due to the time-to-market pressure, only a small number of silicon measurements can be taken at the post-silicon validation stage, especially because the measurements of a number of AMS performance metrics such as bit rate error are extremely time-consuming. Hence, it is impossible to collect a large number of samples for yield estimation for either pre-silicon verification or post-silicon validation.

To address this issue, Bayesian Model Fusion (BMF) has been proposed to accurately estimate the parametric yield and/or the statistical distribution of circuit performance for both pre-silicon verification and post-silicon validation [5-8]. The key idea of BMF is to borrow the knowledge of an early stage (e.g., pre-layout simulation) to accurately estimate the yield and/or distribution at the late stage (e.g., post-layout simulation) with very few late-stage samples. Because the simulation and/or measurement data from the early and late stages are derived from the same circuit, they are expected to be highly correlated. By fusing the early-stage and late-stage data, the parametric yield and/or performance distribution of the late stage can be accurately estimated with few samples and, hence, low computational and/or measurement cost.

The BMF method was previously developed for moment estimation of AMS circuits where only a single performance metric is considered [5-8]. However, the parametric yield value of an AMS circuit is often defined by multiple correlated performance metrics. Motivated by this observation, we propose a novel multivariate moment estimation method in this paper. Particularly, we assume that the probability distribution of multiple AMS performance metrics is jointly (or, multivariate) Gaussian and our objective is to accurately estimate its means vector and covariance matrix.

Towards this goal, we borrow the early-stage data and encode the prior knowledge as a normal-Wishart distribution according to the conjugate prior theory from the statistics community [14]. Next, the prior knowledge is combined with very few late-stage data via Bayesian inference to accurately estimate the mean vector and covariance matrix for multiple AMS performance metrics. As will be demonstrated by our experimental results in Section 5, the proposed method can achieve up to 16× cost reduction over the traditional Maximum Likelihood Estimation (MLE) method without surrendering any accuracy.

It is important to note that the probability distribution of AMS performance metrics may not be accurately modeled as a jointly Gaussian distribution [9-13]. However, accurately estimating a non-Gaussian distribution often needs to collect a large number of samples and, hence, is not feasible with limited data. Since we focus on the problem of pre-silicon verification and post-silicon validation with an extremely small data set, we constrain our discussions to Gaussian distribution in this paper. How to extend the proposed BMF method to other non-Gaussian distributions will be further studied in our future researches (e.g., by estimating and matching the high-order moments).

The reminder of this paper is organized as follows. In Section

---

* Corresponding authors: {yangfan, xzeng}@fudan.edu.cn.

DAC '15, June 07 - 11, 2015, San Francisco, CA, USA

http://dx.doi.org/10.1145/2744769.2744832

2, we review the background of BMF and then derive our proposed method for multivariate moment estimation in Section 3. Several implementation details are further discussed in Section 4. The efficacy of the proposed method is demonstrated by two circuit examples in Section 5. Finally, we conclude in Section 6.

2. BACKGROUND

Several BMF methods have been proposed to estimate the statistics (e.g., moment, distribution) of single performance metrics. In this section, we briefly review these BMF methods.

BMF method has been applied to improve the estimation accuracy of mean and variance with extremely small sample size [7]. Based on the fact that the simulation and measurement data under different circuit configurations and corners are strongly correlated, the authors exploit the correlation information by Bayesian inference to improve the estimation accuracy. Besides the moment estimation problem for multiple populations, the proposed BMF method can also be used to accurately estimate the mean and variance by combining the prior information from the early stage with very few random samples at the late stage.

In practice, the variability of a performance of interest (say, x) can often be approximated as a Gaussian distribution [7]

$$x \sim Gauss\left(\mu,\sigma^2\right), \tag{1}$$

which implies that the performance distribution can be fully specified by its mean μ and variance σ^2.

Next, we will take the problem of mean estimation as an example to illustrate the basic BMF framework proposed in [7]. Traditional MLE method estimates the mean value μ by the sample mean, which is likely to be inaccurate when the sample size is extremely limited. BMF attempts to exploit the correlation between the early-stage and late-stage data to further improve the estimation accuracy. Generally, the unknown mean value μ is assumed to follow a *prior distribution* of the form:

$$\mu \sim p(\mu). \tag{2}$$

Given a value of μ, the *likelihood function* measures the probability of observing specific samples:

$$p(\mathbf{x}|\mu), \tag{3}$$

where $\mathbf{x} = [x^{(1)}\ x^{(2)} \ldots x^{(N)}]^T$ denotes N observed samples. With the prior distribution $p(\mu)$ in (2) and the likelihood function $p(\mathbf{x} \mid \mu)$ in (3), *maximum-a-posteriori* (MAP) estimation is applied to find the mean value μ that is most likely to occur for the prior distribution $p(\mu)$ and the observed data $\mathbf{x}$. It searches for the optimal mean value μ by maximizing the *posterior distribution*:

$$\max_{\mu} \quad p(\mu \mid \mathbf{x}) \propto p(\mu) \cdot p(\mathbf{x} \mid \mu). \tag{4}$$

BMF has also been extended to estimate a non-Gaussian or even discrete distribution. A BMF-BD (Bayesian Model Fusion on Bernoulli Distribution) method is proposed for efficient yield estimation in [5]. BMF-BD is particularly developed to handle the cases where the pre-silicon simulation and/or post-silicon measurement results are binary: either "pass" or "fail", which can be modeled as Bernoulli distribution.

In addition, BMF has been used to re-use the schematic-level simulation result to estimate the post-layout performance distribution in [8]. Namely, for a specific performance of interest, BMF can accurately estimate its late-stage probability density function (PDF) with limited late-stage samples. Please refer to [8] for the detail of the method.

The aforementioned BMF methods can accurately estimate the statistics of single late-stage performance metric with few late-stage samples. However, for most AMS circuits, the parametric yield values are defined by multiple correlated performance metrics. The marginal statistics of single performance obtained by the traditional BMF methods is not enough for circuit optimization and yield estimation. It, in turn, motivates us to develop BMF method for multiple correlated performance metrics.

3. PROPOSED APPROACH

In this section, we develop our proposed BMF method to estimate the multivariate moments for multiple correlated performance metrics. We first briefly describe the problem formulation, and then derive the corresponding Bayesian inference for multivariate moment estimation.

3.1 Problem Formulation

We define a d-dimensional random vector $\mathbf{X} = [x_1, x_2, \ldots, x_d]$ to model d performance metrics where x_i denotes the i-th metric. These performance metrics are all influenced by many process parameters to different extents.

The probability distribution of AMS performance metrics may not be precisely modeled as a jointly Gaussian distribution. However, with limited samples, it is impossible to accurately estimate the non-Gaussian distributions of the correlated performance metrics. Therefore, we assume the joint distribution of the multiple performance metrics to be Gaussian and attempt to estimate the mean vector and covariance matrix of the multiple performance metrics. The jointly Gaussian distribution can approximate non-Gaussian distribution with acceptable accuracy and the covariance matrix of the performance metrics also provides useful guides for design optimization and yield improvement.

We assume the jointly Gaussian distribution of $\mathbf{X}$ denoted by

$$\mathbf{X} \sim N_d\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right), \tag{5}$$

with d-dimensional mean vector

$$\boldsymbol{\mu} = [E(x_1), E(x_2), \cdots, E(x_d)], \tag{6}$$

and $d \times d$ covariance matrix

$$\boldsymbol{\Sigma} = [\mathrm{Cov}\left(x_i, x_j\right)],\ i = 1,2,\ldots,d;\ j = 1,2,\ldots,d \,. \tag{7}$$

The density function of the distribution can be expressed as

$$p\left(\mathbf{X}\middle|\boldsymbol{\mu}, \boldsymbol{\Sigma}\right) = \left(2\pi\right)^{-d/2}\left|\boldsymbol{\Sigma}\right|^{-1/2}\exp\left(-\frac{1}{2}\left(\mathbf{X}-\boldsymbol{\mu}\right)^T\boldsymbol{\Sigma}^{-1}\left(\mathbf{X}-\boldsymbol{\mu}\right)\right). \tag{8}$$

It is fully characterized by the first two moments: mean vector and covariance matrix. Hence, the goal of our work is to estimate the mean vector and covariance matrix of the late-stage distribution with the help of early-stage knowledge and few late-stage samples.

After observing n data samples, the likelihood function can be expressed as:

$$p\left(\mathbf{D}\middle|\boldsymbol{\mu},\boldsymbol{\Sigma}\right) = \left(2\pi\right)^{-nd/2}\left|\boldsymbol{\Sigma}\right|^{-n/2}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(\mathbf{X}_i-\boldsymbol{\mu}\right)^T\boldsymbol{\Sigma}^{-1}\left(\mathbf{X}_i-\boldsymbol{\mu}\right)\right), \tag{9}$$

where $\mathbf{D} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n]$ denotes the observed samples.

Traditional MLE method selects the value set of model parameters that maximizes the likelihood function. Thus, the mean vector and covariance matrix are estimated as followings:

$$\boldsymbol{\mu}_{MLE} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i, \tag{10}$$

$$\boldsymbol{\Sigma}_{MLE} = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{X}_i-\boldsymbol{\mu}_{MLE}\right)\left(\mathbf{X}_i-\boldsymbol{\mu}_{MLE}\right)^T. \tag{11}$$

Studying (10) and (11) reveals that the MLE method reaches the estimated results only based on late-stage data samples.

3.2 Prior Knowledge Definition

For simplicity, we use the precision matrix $\mathbf{\Lambda} = \mathbf{\Sigma}^{-1}$ instead of covariance matrix $\mathbf{\Sigma}$ in the following analysis. We borrow the theory of conjugate prior [14] from the statistics community. If a prior distribution and a posterior distribution are in the same family (i.e., have the same function form), they are called conjugate distributions. The prior distribution is then called a conjugate prior for the likelihood function and leads to a greatly simplified Bayesian analysis. The conjugate prior of jointly Gaussian distribution is normal-Wishart distribution [14]. Thus, we use normal-Wishart distribution $p(\boldsymbol{\mu}, \mathbf{\Lambda}|\ \boldsymbol{\mu}_0, v_0, \kappa_0, \mathbf{T}_0)$ to encode the prior knowledge and estimate the moments of late-stage distributions. Normal-Wishart distribution can be expressed as:

$$
\begin{aligned}
&p\left(\boldsymbol{\mu}, \mathbf{\Lambda} \mid \boldsymbol{\mu}_0, \kappa_0, v_0, \mathbf{T}_0\right) = N_d\left(\boldsymbol{\mu} \middle| \boldsymbol{\mu}_0, \left(\kappa_0 \mathbf{\Lambda}\right)^{-1}\right) Wi_{v_0}\left(\mathbf{\Lambda} \middle| \mathbf{T}_0\right) \\
&= \frac{1}{Z_0} \left|\mathbf{\Lambda}\right|^{\frac{1}{2}} \exp\left(-\frac{\kappa_0}{2}\left(\boldsymbol{\mu}-\boldsymbol{\mu}_0\right)^T \mathbf{\Lambda}\left(\boldsymbol{\mu}-\boldsymbol{\mu}_0\right)\right) \\
&\cdot \left|\mathbf{\Lambda}\right|^{(v_0-d-1)/2} \exp\left(-\frac{1}{2} tr\left(\mathbf{T}_0^{-1} \mathbf{\Lambda}\right)\right)
\end{aligned}
, \tag{12}
$$

where $\Gamma_d(\bullet)$ denotes the d-dimensional multivariate gamma function and $tr(\bullet)$ is the trace of matrix. Z_0 is a normalization coefficient as:

$$
Z_0 = \left(\frac{2\pi}{\kappa_0}\right)^{d/2} \left|\mathbf{T}_0\right|^{v_0/2} 2^{v_0 d/2} \Gamma_d\left(v_0/2\right). \tag{13}
$$

Normal-Wishart distribution is a combination of Gaussian distribution and Wishart distribution. $\boldsymbol{\mu}_0$, v_0, κ_0 and $\mathbf{T}_0$ are four hyper-parameters that control the shape of the distribution. Vector $\boldsymbol{\mu}_0$ and scalar κ_0 influence the location and the shape of the Gaussian component. Scalar $v_0 \geq d$ and the $d \times d$ matrix $\mathbf{T}_0$ are the degrees of freedom and the scale matrix of the Wishart component, respectively.

We define $\boldsymbol{\mu}_\mathbf{M}$ and $\mathbf{\Lambda}_\mathbf{M}$ as the mode of the two parameters $\boldsymbol{\mu}$ and $\mathbf{\Lambda}$ in (12) at which $p(\boldsymbol{\mu}, \mathbf{\Lambda})$ is peaked. In order to maximize $p(\boldsymbol{\mu}, \mathbf{\Lambda})$, we consider $\ln p(\boldsymbol{\mu}, \mathbf{\Lambda})$ as follows.

$$
\begin{aligned}
&\ln p\left(\boldsymbol{\mu}, \mathbf{\Lambda} \middle| \boldsymbol{\mu}_0, \kappa_0, v_0, \mathbf{T}_0\right) \\
&\sim -\kappa_0\left(\boldsymbol{\mu}-\boldsymbol{\mu}_0\right)^T \mathbf{\Lambda}\left(\boldsymbol{\mu}-\boldsymbol{\mu}_0\right) + \left(v_0-d\right)\ln\left(\left|\mathbf{\Lambda}\right|\right) - tr\left(\mathbf{T}_0^{-1}\mathbf{\Lambda}\right)
\end{aligned}
. \tag{14}
$$

By taking the partial derivatives of $\ln p(\boldsymbol{\mu}, \mathbf{\Lambda})$ with regard to $\boldsymbol{\mu}$ and $\mathbf{\Lambda}$ respectively and then setting them to zero, we have:

$$
\boldsymbol{\mu}_\mathbf{M} = \boldsymbol{\mu}_0, \tag{15}
$$

$$
\mathbf{\Lambda}_\mathbf{M} = \left(v_0 - d\right)\mathbf{T}_\mathbf{0}. \tag{16}
$$

We then set up the following constraints for the hyper-parameters:

$$
\boldsymbol{\mu}_\mathbf{M} = \boldsymbol{\mu}_0 = \boldsymbol{\mu}_\mathbf{E}, \tag{17}
$$

$$
\mathbf{\Lambda}_\mathbf{M} = \left(v_0 - d\right)\mathbf{T}_\mathbf{0} = \mathbf{\Lambda}_\mathbf{E}, \tag{18}
$$

where $\boldsymbol{\mu}_\mathbf{E}$ denotes mean vector of early stage and $\mathbf{\Lambda}_\mathbf{E}$ denotes corresponding precision matrix. The prior distribution defined by (12), (17) and (18) implies that the late-stage mean vector and covariance matrix are likely to be similar to the early-stage mean vector and covariance matrix, because $p(\boldsymbol{\mu}, \mathbf{\Lambda})$ is peaked at $(\boldsymbol{\mu}_\mathbf{M}, \mathbf{\Lambda}_\mathbf{M})$.

From (17) and (18) we obtain

$$
\boldsymbol{\mu}_0 = \boldsymbol{\mu}_\mathbf{E}, \tag{19}
$$

$$
\mathbf{T}_\mathbf{0} = \frac{1}{v_0 - d}\mathbf{\Lambda}_\mathbf{E}. \tag{20}
$$

Substituting (19) and (20) into (12) yields:

$$
\begin{aligned}
&p\left(\boldsymbol{\mu}, \mathbf{\Lambda} \mid \kappa_0, v_0\right) = \frac{1}{Z_0}\left|\mathbf{\Lambda}\right|^{\frac{1}{2}} \exp\left(-\frac{\kappa_0}{2}\left(\boldsymbol{\mu}-\boldsymbol{\mu}_\mathbf{E}\right)^T \mathbf{\Lambda}\left(\boldsymbol{\mu}-\boldsymbol{\mu}_\mathbf{E}\right)\right) \\
&\cdot \left|\mathbf{\Lambda}\right|^{(v_0-d-1)/2} \exp\left(-\frac{1}{2} tr\left(\left(v_0 - \mathrm{d}\right)\mathbf{\Lambda}_\mathbf{E}^{-1}\mathbf{\Lambda}\right)\right)
\end{aligned}
, \tag{21}
$$

where only two hyper-parameters v_0, κ_0 should be determined. We estimate the optimal value of v_0, κ_0 by a two-dimensional cross validation process, as will be described in Section 4.2.

3.3 Maximum-A-Posteriori Estimation

We combine the prior distribution (21) with few late-stage samples to estimate the moments of late-stage distribution. Maximum-A-Posteriori Estimation (MAP) is used in this procedure. Since it takes the information of late-stage samples into account, the prior distribution uncertainty can be greatly reduced.

Using the precision matrix $\mathbf{\Lambda} = \mathbf{\Sigma}^{-1}$ instead of covariance matrix $\mathbf{\Sigma}$, the likelihood function (9) can be rewritten as

$$
p\left(\mathbf{D} \mid \boldsymbol{\mu}, \mathbf{\Lambda}\right) = \left(2\pi\right)^{-nd/2}\left|\mathbf{\Lambda}\right|^{n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(\mathbf{X}_i - \boldsymbol{\mu}\right)^T \mathbf{\Lambda}\left(\mathbf{X}_i - \boldsymbol{\mu}\right)\right). \tag{22}
$$

After obtaining n late-stage samples $\mathbf{D} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n]$, which follows a multivariate Gaussian distribution, we can get the posterior distribution in the light of Bayes' theorem. The posterior distribution $p(\boldsymbol{\mu}, \mathbf{\Lambda}|\ \mathbf{D})$ is proportional to the prior distribution $p(\boldsymbol{\mu}, \mathbf{\Lambda}|\ v_0, \kappa_0)$ in (21) multiplied by the likelihood function $p(\mathbf{D}|\ \boldsymbol{\mu}, \mathbf{\Lambda})$ in (22):

$$
\begin{aligned}
&p\left(\boldsymbol{\mu}, \mathbf{\Lambda} \mid \mathbf{D}\right) \propto \left|\mathbf{\Lambda}\right|^{\frac{1}{2}} \exp\left(-\frac{\kappa_n}{2}\left(\boldsymbol{\mu}-\boldsymbol{\mu}_\mathbf{n}\right)^T \mathbf{\Lambda}\left(\boldsymbol{\mu}-\boldsymbol{\mu}_\mathbf{n}\right)\right) \\
&\cdot \left|\mathbf{\Lambda}\right|^{(v_n-d-1)/2} \exp\left(-\frac{1}{2} tr\left(\mathbf{T}_\mathbf{n}^{-1}\mathbf{\Lambda}\right)\right)
\end{aligned}
, \tag{23}
$$

where

$$
\boldsymbol{\mu}_\mathbf{n} = \frac{\kappa_0 \boldsymbol{\mu}_\mathbf{E} + n\bar{\mathbf{X}}}{\kappa_0 + n}, \tag{24}
$$

$$
\mathbf{T}_\mathbf{n}^{-1} = \left(v_0 - d\right)\mathbf{\Lambda}_\mathbf{E}^{-1} + \mathbf{S} + \frac{\kappa_0 n}{\kappa_0 + n}\left(\boldsymbol{\mu}_\mathbf{E} - \bar{\mathbf{X}}\right)\left(\boldsymbol{\mu}_\mathbf{E} - \bar{\mathbf{X}}\right)^T, \tag{25}
$$

$$
\mathbf{S} = \sum_{i=1}^{n}\left(\mathbf{X}_i - \bar{\mathbf{X}}\right)\left(\mathbf{X}_i - \bar{\mathbf{X}}\right)^T, \tag{26}
$$

$$
v_n = v_0 + n, \tag{27}
$$

$$
\kappa_n = \kappa_0 + n, \tag{28}
$$

and $\bar{\mathbf{X}}$ is the mean vector of the n late-stage samples.

Note that the posterior distribution $p(\boldsymbol{\mu}, \mathbf{\Lambda}|\ \mathbf{D})$ models the uncertainty of $\boldsymbol{\mu}$ and $\mathbf{\Lambda}$ after given the n late-stage data samples. Our goal is to find the optimal value of $\boldsymbol{\mu}$ and $\mathbf{\Lambda}$ that are most likely to occur according to the observed late-stage samples. Thus we apply the MAP to find the values of $\boldsymbol{\mu}_{MAP}$ and $\mathbf{\Lambda}_{MAP}$ that make the posterior distribution reach its maximum, i.e., the mode of the posterior distribution $p(\boldsymbol{\mu}, \mathbf{\Lambda}|\ \mathbf{D})$. Since the posterior distribution in (23) has the same form as the prior distribution in (12), the values of $\boldsymbol{\mu}_{MAP}$ and $\mathbf{\Lambda}_{MAP}$ are easy to determine similarly:

$$
\boldsymbol{\mu}_{MAP} = \boldsymbol{\mu}_\mathbf{n}, \tag{29}
$$

$$
\mathbf{\Lambda}_{MAP} = \left(v_n - d\right)\mathbf{T}_\mathbf{n}. \tag{30}
$$

By substituting (24) and (25) into (29) and (30) respectively and replacing $\mathbf{\Lambda}_{MAP}$ and $\mathbf{\Lambda}_{\mathbf{E}}$ with $\mathbf{\Sigma}_{MAP}$ and $\mathbf{\Sigma}_{\mathbf{E}}$, we can derive the MAP estimation of the moments as:

$$\boldsymbol{\mu}_{MAP}=\frac{\kappa_0\boldsymbol{\mu}_{\mathbf{E}}+n\overline{\mathbf{X}}}{\kappa_0+n}, \tag{31}$$

$$\mathbf{\Sigma}_{MAP}=\frac{(v_0-d)\mathbf{\Sigma}_{\mathbf{E}}+\mathbf{S}+\dfrac{\kappa_0 n}{\kappa_0+n}\left(\boldsymbol{\mu}_{\mathbf{E}}-\overline{\mathbf{X}}\right)\left(\boldsymbol{\mu}_{\mathbf{E}}-\overline{\mathbf{X}}\right)^T}{v_0+n-d}. \tag{32}$$

Eq. (31) and (32) show the important roles of hyper-parameters v_0, κ_0 in controlling $\boldsymbol{\mu}_{MAP}$ and $\mathbf{\Sigma}_{MAP}$ during the balance between early-stage and late-stage information. For $\boldsymbol{\mu}_{MAP}$ (controlled by κ_0), in one extreme case that κ_0 is sufficiently large, we have:

$$\boldsymbol{\mu}_{MAP}\approx\boldsymbol{\mu}_{\mathbf{E}}, \tag{33}$$

which reveals that the prior knowledge is already accurate enough. In the other extreme case that κ_0 is sufficiently small, we have:

$$\boldsymbol{\mu}_{MAP}\approx\overline{\mathbf{X}}=\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i, \tag{34}$$

which is nearly identical to the MLE result in (10). This means that the prior knowledge is incorrect and the Bayesian inference almost neglects it. Similarly, for $\mathbf{\Sigma}_{MAP}$ (mainly controlled by v_0), in the case that v_0 are sufficient large, we have:

$$\mathbf{\Sigma}_{MAP}\approx\mathbf{\Sigma}_{\mathbf{E}}; \tag{35}$$

In contrast, in the case that v_0 is sufficiently small (close to d), the numerator is dominated by $\mathbf{S}$, and $\mathbf{\Sigma}_{MAP}$ is largely dependent on the late-stage knowledge. In the special case that $\kappa_0 = 0$ and $v_0 = d$, (32) is reduced to MLE estimates:

$$\mathbf{\Sigma}_{MAP}=\frac{\mathbf{S}}{n}=\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{X}_i-\overline{\mathbf{X}}\right)\left(\mathbf{X}_i-\overline{\mathbf{X}}\right)^T. \tag{36}$$

From the aforementioned discussion, we can interpret the hyper parameters v_0 and κ_0 as the indicators of credibility on early-stage knowledge. Hence, they should be appropriately set before MAP estimation. We will present a two-dimensional cross validation strategy in Section 4.2 to ensure this requirement.

3.4 Summary

Algorithm 1 summarizes the major steps of our proposed BMF method for multivariate moment estimation. Starting from two given early-stage distribution moments $\boldsymbol{\mu}_{\mathbf{E}}$ and $\mathbf{\Lambda}_{\mathbf{E}}$ as the prior knowledge and a set of late-stage samples $\mathbf{D} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n]$, we first calculate the mean vector of the n observed sample. Then we search for the optimal hyper-parameter v_0, κ_0 for prior normal-Wishart $p(\boldsymbol{\mu},\mathbf{\Lambda}|v_0,\kappa_0)$ in (21) by a two-dimensional cross validation discussed in Section 4.2. Once the prior distribution is fully specified with the optimal v_0 and κ_0 values, we combine the prior distribution $p(\boldsymbol{\mu}, \mathbf{\Lambda}|v_0, \kappa_0)$ with the late-stage samples $\mathbf{D}$ to estimate the late-stage distribution moments $\boldsymbol{\mu}_{MAP}$ and $\mathbf{\Sigma}_{MAP}$ in (31) and (32) based on MAP. Some important implementation details will be described in Section 4.

Algorithm 1:BMF for Multivariate moment Estimation

1. Start from two given early-stage distribution moments $\boldsymbol{\mu}_{\mathbf{E}}$ and $\mathbf{\Lambda}_{\mathbf{E}}$ and a set of late-stage samples $\mathbf{D} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n]$.
2. Calculate the mean vector of the n observed late-stage samples $\overline{\mathbf{X}}$.
3. Apply two-dimensional cross validation strategy to find the optimal hyper-parameters v_0, κ_0.
4. Determine the late-stage distribution moments $\boldsymbol{\mu}_{MAP}$ and $\mathbf{\Sigma}_{MAP}$ in (31) and (32) based on MAP.

4. IMPLEMENTATION DETAILS

To make the proposed BMF framework practically efficient, we also need to consider several implementation issues. In this section, we will discuss these implementation details, including (i) performance shift and scaling, and (ii) two-dimensional cross validation.

4.1 Performance Shift and Scaling

In the proposed BMF method, we assume that the early-stage and late-stage performance distributions pdf_E and pdf_L are quite similar. Unfortunately, in many practical applications, although the shapes of early-stage and late-stage distributions are similar, the nominal performance values do not match. For example, in the schematic-level and post-layout simulations of an AMS circuit, the measured performance metrics (e.g., gain, bandwidth, power, etc.) may be significantly different due to interconnect parasitic parameters. Directly applying BMF estimation will thus lead to inaccurate result. To optimize the prior knowledge and observed samples, we propose a shift and scaling strategy.

First we measure the nominal performance values $\mathbf{P}_{E,NOM}$ and $\mathbf{P}_{L,NOM}$ by running the early-stage and late-stage simulation once respectively. $\mathbf{P}_{E,NOM}$ and $\mathbf{P}_{L,NOM}$ are actually a set of nominal values in regard to multiple performances in the early stage and late stage. Then we shift the early-stage and late-stage distributions pdf_E and pdf_L by $\mathbf{P}_{E,NOM}$ and $\mathbf{P}_{L,NOM}$ respectively.

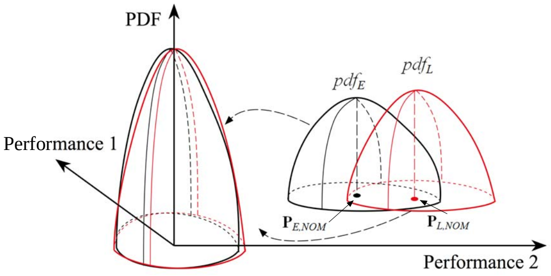


Figure 1. The early-stage and late-stage multi-performance distributions pdf_E and pdf_L are shifted and scaled in each performance dimension so that they become "isotropic" and similar, before the BMF method is applied.

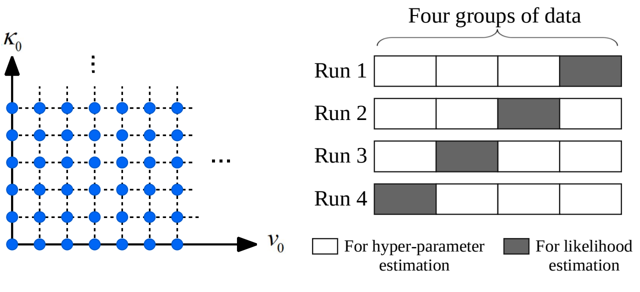
In addition, we also perform a performance scaling. When dealing with multiple performances, we note that the magnitude of different performance metrics vary widely. Taking an operational amplifier for example, the gain and power metrics may differ by more than seven orders of magnitude. This increases the risk of numeric problems in calculation. On top of that, the estimation errors of performance metrics with small values are likely to be covered up, even if they are relatively significant in their own dimension. So after the shift operation, we further scale both stages' data by the standard deviation of early-stage in each dimension. Since the standard deviation of both stages are supposed to be similar, we can still keep the shifted distributions in analogical shapes after scaling. Figure 1 shows an example of the joint distributions of two performance metrics. After the aforementioned performance shift and scaling, the early-stage and late-stage distributions are transformed to two origin-centered "isotropic" ones, which means they have near-zero mean value and near-one standard deviation in each performance dimension.

4.2 Two-dimensional Cross Validation

As analyzed in Section 3.3, the values of v_0 and κ_0 control the confidence of prior knowledge and therefore should be properly set. Similar to [8], we employ the idea of cross validation from the statistics community [14] to automatically find the optimal v_0 and κ_0 based on few late-stage samples.

To achieve this goal, we run the BMF flow repeatedly with different sets of v_0 and κ_0 to estimate the late-stage distribution moments and monitor the estimation accuracy. Figure 2(a) shows the search space of v_0 and κ_0. All combinations (represented as the points) within the pre-defined range are chosen as candidates. Note that the accuracy is expected to vary as the hyper-parameter v_0 and κ_0 change. The optimal v_0 and κ_0 are determined as the corresponding set of values when the maximum accuracy is achieved.

To assess the estimation accuracy of a given set of the hyper-parameter v_0 and κ_0, we apply a Q-fold cross validation strategy. In the Q-fold cross validation process, we divide the given post-stage samples into Q groups and run the BMF algorithm (i.e. Algorithm 1) for Q times. At each run, a training set consisting of Q-1 groups is used to estimate the moments of late-stage distribution. After that, a testing set consisting of the remaining group is used to assess the estimation accuracy with respect to the chosen hyper-parameters based on the likelihood function (9). Larger likelihood function value indicates more accurate estimation. Then we change the training set and the testing set and repeat the above steps in a new run. Each run results in a likelihood value that is measured from a unique group of samples. After a complete cross-validation process of Q runs, we then calculate the average of the Q likelihood values and use it to indicate the estimation accuracy of the given hyper-parameters v_0 and κ_0. In this process, over-fitting can be easily detected, since we use separate training set and testing set in each run. A simple example for a four-fold cross-validation process was illustrated in Figure 2(b).



(a) (b)

Figure 2. (a) Illustration of the two-dimensional search space of v_0 and κ_0. (b) Illustration of a four-fold cross validation process for each combination of v_0 and κ_0.

5. NUMERICAL EXAMPLES

In this section, we use two circuit examples to demonstrate the efficiency of our proposed BMF method. For testing and comparison purpose, we implement two different multivariate moment estimation method: (i) the conventional MLE method, and (ii) the proposed BMF method. All experiments are performed on a server with 2.5GHz dual-core CPU and 16GB memory.

5.1 Operational Amplifier

In this example, we use a two-stage operational amplifier designed in a 45nm CMOS process. A simplified schematic of the amplifier is shown in Figure 3. We consider the schematic-level design as early stage and the post-layout design as late stage. Five correlated performance metrics including gain, -3dB bandwidth, power, offset and phase margin are measured in both stages. We aim to accurately estimate the mean vector and covariance matrix of the five performance variables of late stage by borrowing the prior knowledge from the early stage.

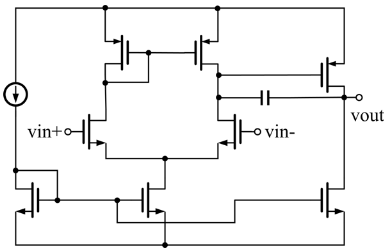


Figure 3. A simplified schematic is shown for a two-stage operational amplifier in a 45nm CMOS process.

For testing and comparison purpose, we generate 5000 Monte-Carlo samples by both schematic-level and post-layout simulations, in which the device-level variations of all transistors are considered.

To compare the accuracy between MLE and our BMF method, we first define the estimation error of the mean vector and covariance matrix respectively. Since we are dealing with multiple variables, we need a proper error criterion to equally reflect the relative error in every dimension. We calculate the error based on the shifted and scaled data. As mentioned in Section 4.1, after the shift and scaling operation, the order of magnitudes in each performance dimension are similar. Due to this “isotropic” property, the estimation errors in every dimension are taken into consideration equally so that the error of small value performance will not be concealed. Hence, the overall estimation errors of the multivariate moments are defined as:

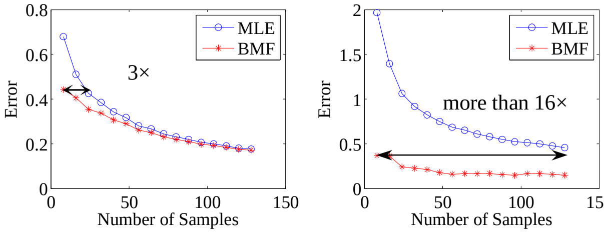
$$Error\_mean = \left\| \boldsymbol{\mu}_{ESTI} - \boldsymbol{\mu}_{EXACT} \right\|_2 , \tag{37}$$

$$Error\_cov = \left\| \boldsymbol{\Sigma}_{ESTI} - \boldsymbol{\Sigma}_{EXACT} \right\|_F , \tag{38}$$

where $\boldsymbol{\mu}_{ESTI}$ denotes the estimated mean vector and $\boldsymbol{\Sigma}_{ESTI}$ denotes the estimated covariance matrix; $\boldsymbol{\mu}_{EXACT}$ denotes the exact mean vector and $\boldsymbol{\Sigma}_{EXACT}$ denotes the exact covariance matrix; $\|\bullet\|_2$ denotes the 2-norm of vector and $\|\bullet\|_F$ denotes the Frobenius-norm of matrix. We use the absolute error rather than relative error because the shift and scaling operation is actually a normalization procedure in a sense. In this case, the absolute error reflects the relative mismatch between the estimated and actual distribution shapes.

To include two extreme situations in Section 3.3, we search the optimal values of the hyper parameters v_0 and κ_0 from 1 to 1000 respectively. Figure 4(a) and Figure 4(b) show the estimation errors of late-stage mean vector and covariance matrix respectively, as a function of the number of late-stage samples. The errors are calculated from 100 repeated runs based on independent samples to average out random fluctuations. Note that the BMF method achieves high accuracy in covariance matrix estimation, even when the sample number is less than 20. However, MLE require more than 128 points to achieve the same accuracy. Studying the plot reveals that BMF achieves more than

16× cost reduction over MLE in covariance matrix estimation. For mean vector estimation, BMF also achieves nearly 3× cost reduction when the sample number is extremely small. On the other hand, we find the optimized values of κ_0 are quite small (e.g., 4.67 when post-stage sample number is 32) while the optimized values of v_0 are significantly larger (e.g., 557.3 when post-stage sample number is 32). According to the analysis in Section 3.3, these optimized values of hyper parameters imply that the early-stage knowledge of mean vector is not accurate enough thus less useful for reference than the early-stage knowledge of covariance matrix. As a result, the BMF method relies largely on post-stage samples when estimating the mean vector and this explains why the cost reduction in Figure 4(a) is not as significant as in Figure 4(b).

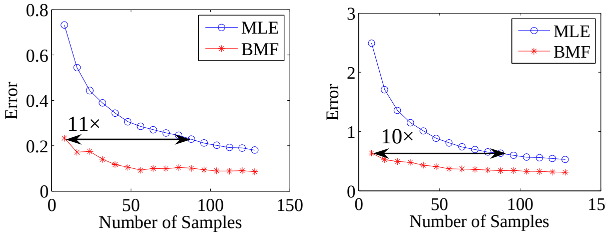


(a) (b)

Figure 4. (a) The estimation error of late-stage mean vector is plotted as a function of the number of late-stage samples. (b) The estimation error of late-stage covariance matrix is plotted as a function of the number of late-stage samples.

5.2 Analog to Digital Converter

In this example, we consider a flash analog to digital converter in a 0.18μm CMOS process. Five correlated performance metrics including signal-to-noise ratio (SNR), signal-to-noise-and-distribution ratio (SINAD), spurious-free-dynamic-range (SFDR), total-harmonic-distortion (THD) and power are measured in the schematic-level and the post-layout design stages. For testing and comparison purpose, 1000 random samples are generated by both schematic-level and post-layout simulations.



(a) (b)

Figure 5. (a) The estimation error of late-stage mean vector is plotted as a function of the number of late-stage samples. (b) The estimation error of late-stage covariance matrix is plotted as a function of the number of late-stage samples.

Figure 5(a) and Figure 5(b) plot the estimation errors of late-stage mean vector and covariance matrix, respectively, as a function of the number of late-stage samples. BMF provides superior accuracy for not only the mean vector but also the covariance matrix estimation. In both plots, even if the number of late-stage samples is as small as eight, the error of BMF is already small enough while MLE requires more than 10× samples to achieve the same accuracy. In this example, the optimized values of v_0 and κ_0 are all relatively large (e.g., 521.9 for κ_0 and 558.8 for v_0 when post-stage sample number is 32). This implies that the early-stage of mean vector and covariance matrix are both very useful for reference and that is why we achieve more than 10× speed-up in both estimations.

6. CONCLUSIONS

In this paper, we propose a Bayesian model fusion method for efficient multivariate moment estimation of multiple correlated performance metrics by borrowing the prior knowledge from the early stage. The multiple performance metrics are modeled as a jointly Gaussian distribution and the prior knowledge is encoded as a normal-Wishart distribution. The late-stage multivariate moments are accurately estimated by Bayesian inference with very few late stage samples. Our circuit examples demonstrate that the proposed method can achieve up to 16× cost reduction over the traditional MLE method without surrendering any accuracy.

7. ACKNOWLEDGEMENTS

This work was supported in part by the National Natural Science Foundation of China under Project 61125401, Project 61376040, Project 61274032, Project 91330201, and Project 61474026, in part by the National Basic Research Program of China under Grant 2011CB309701, in part by the National Major Science and Technology Special Project of China under Grant 2011ZX01035-001-001-003 and Grant 2014ZX02301002-002, in part by the Shanghai Science and Technology Committee under Project 13XD1401100, and in part by the Chen Guang Project through Shanghai Municipal Education Commission and Shanghai Education Development Foundation.

8. REFERENCES

[1] P. Desrumaux, et al., “An efficient control variates method for yield estimation of analog circuits based on a local model,” *IEEE ICCAD*, pp. 415-421, 2012.

[2] S. Mitra, et al., “Post-silicon validation opportunities, challenges and recent advances,” *IEEE DAC*, pp. 12-17, 2010.

[3] X. Li, “Post-silicon performance modeling and tuning of analog/mixed-signal circuits via Bayesian model fusion,” *IEEE ICCAD*, pp. 551-552, 2012.

[4] S. Sun, et al., “Indirect performance sensing for on-chip self-healing of analog and RF circuits,” *IEEE Trans. on CAS-I*, vol. 61, no. 8, pp. 2243-2252, Aug. 2014.

[5] C. Fang, et al., “BMF-BD: Bayesian model fusion on Bernoulli distribution for efficient yield estimation of integrated circuits,” *IEEE DAC*, 2014.

[6] 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.

[7] C. Gu, et al., “Efficient moment estimation with extremely small sample size via Bayesian inference for analog/mixed-signal validation,” *IEEE DAC*, 2013.

[8] X. Li, et al., “Efficient parametric yield estimation of analog/mixed-signal circuits via Bayesian model fusion,” *IEEE ICCAD*, pp. 627-634, 2012.

[9] X. Li, et al., “Efficient parametric yield extraction for multiple correlated non-Normal performance distributions of analog/RF circuits,” *IEEE DAC*, pp. 928-933, 2007.

[10] X. Li, et al., “Asymptotic probability extraction for non-Normal distributions of circuit performance,” *IEEE ICCAD*, pp. 2-9, 2004.

[11] A. Dharchoudhury, et al., “Worse-case analysis and optimization of VLSI circuit performance,” *IEEE Trans. CAD*, vol. 14, no. 4, pp. 481-492, Apr. 1995.

[12] E. Felt, et al., “Hierarchical statistical characterization of mixed-signal circuits using behavioral modeling,” *IEEE ICCAD*, pp. 374-380, 1996.

[13] J. Swidainski, et al., “Statistical behavioral modeling of integrated circuits,” *IEEE ISCAS*, vol. 6, pp. 98-101, 1998.

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