BMF-BD: Bayesian Model Fusion on Bernoulli Distribution for Efficient Yield Estimation of Integrated Circuits

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
Accurate yield estimation is one of the important yet challenging tasks for both pre-silicon verification and post-silicon validation. In this paper, we propose a novel method of *Bayesian model fusion on Bernoulli distribution* (BMF-BD) for efficient yield estimation at the late stage by borrowing the prior knowledge from an early stage. 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". The key idea is to model the binary simulation/measurement outcome as a Bernoulli distribution and then encode the prior knowledge as a Beta distribution based on the theory of conjugate prior. As such, the late-stage yield can be accurately estimated through Bayesian inference with very few late-stage samples. Several circuit examples demonstrate that BMF-BD achieves up to 10× cost reduction over the conventional estimator without surrendering any accuracy.

1. INTRODUCTION
The aggressive scaling of integrated circuits (ICs) results in large-scale process uncertainties, including parametric variations and catastrophic defects [1]-[2]. Both of them may lead to substantial yield loss at an advanced technology node. Hence, accurate yield estimation has been identified as one of the top priorities for both pre-silicon verification [3]-[5] and post-silicon validation [6]-[7] in order to improve circuit performance and/or reduce manufacturing cost.

Recently, a number of emerging design techniques (e.g., post-silicon tuning [8]-[11]) have been proposed to combat the variability issue and, hence, maintain the continuous scaling of integrated circuits. While adopting these new techniques has been demonstrated with great success, the complexity of today's integrated circuits enormously grows. It, in turn, creates numerous challenges for yield estimation which requires to collect a large number of random samples by either numerical simulation (i.e., for pre-silicon verification) or silicon measurement (i.e., for post-silicon validation).

- *Pre-silicon verification*: Simulating a complex integrated circuit is not trivial. Instead, it can be extremely time-consuming. Therefore, collecting a large number of random samples by Monte Carlo simulation can be challenging, if not impossible.
- *Post-silicon validation*: Given the large complexity of the state-of-the-art integrated circuits, appropriately testing a silicon chip to make a "pass" or "fail" decision is not trivial either. For instance, it can take several minutes or even hours to thoroughly check hundreds of testing items over all environmental corners for a system on chip (SoC). For this reason, only a small number of chips can be affordably tested for post-silicon validation.

To address this cost issue posed by data collection, *Bayesian model fusion* (BMF) has been proposed to accurately estimate the statistics (e.g., distribution, yield, etc.) of circuit performance for both pre-silicon verification [12]-[13] and post-silicon validation [14]-[15]. BMF attempts to borrow the simulation and/or measurement data from an early stage to accurately estimate the statistics at the late stage with very few late-stage samples. For instance, the schematic-level simulation results can be re-used to estimate the post-layout performance distributions [12], and the measurement data from an early tape-out can be borrowed to build the statistical models for the late tape-out [15]. By intelligently minimizing the amount of required simulation and/or measurement data, BMF substantially reduces the verification and/or validation cost at the late stage. A brief summary of BMF can be found in [16].

The conventional BMF technique, however, can efficiently handle continuous performance metrics (e.g., delay of a digital path, gain of an analog amplifier, etc.) only, as it assumes that the underlying performance distribution is continuous [12]-[16]. In many practical applications, the simulation and/or measurement results are often binary: either "pass" or "fail". For instance, testing a silicon chip implemented for a digital system cannot tell the exact delay of its critical path; instead, we only know whether the chip passes or fails the specification. Hence, the conventional BMF method is no longer applicable here.

In this paper, we propose a new method of *Bayesian model fusion on Bernoulli distribution* (BMF-BD) to accommodate the binary outcome of pre-silicon simulation and/or post-silicon measurement. Our key idea is to model the late-stage simulation and/or measurement result as a Bernoulli random variable that takes binary values. Meanwhile, the prior knowledge is extracted from the early-stage data and encoded as a Beta distribution, based on the theory of *conjugate prior* borrowed from the statistics community [17]. Finally, the prior knowledge is combined with very few late-stage data through Bayesian inference to accurately predict the late-stage yield. As will be demonstrated by our circuit examples in Section 4, the proposed BMF-BD method achieves up to 10× cost reduction over the conventional estimator without surrendering any accuracy.

The remainder of this paper is organized as follows. In Section 2, we review the important background on BMF, and then derive our proposed BMF-BD formulation in Section 3. The efficacy of BMF-BD is demonstrated by several examples in Section 4. Finally, we conclude in Section 5.

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

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
DAC '14, June 01 - 05 2014, San Francisco, CA, USA
Copyright 2014 ACM 978-1-4503-2730-5/14/06…$15.00.
http://dx.doi.org/10.1145/2593069.2593099

2. BACKGROUND

Several BMF methods have been developed in the literature to estimate the statistics of continuously distributed circuit performances [12]-[16]. In this section, we briefly review these conventional BMF methods.

2.1 Bayesian Model Fusion for Moment Estimation

An efficient BMF method has been proposed in [14] to improve the estimation accuracy of mean and variance with extremely small sample size. It relies on the assumption that the simulation and measurement data collected for different populations (i.e., different circuit configurations and/or corners) are strongly correlated. The correlation information can thus be exploited by a Bayesian framework to improve the estimation accuracy. Although the authors of [14] mainly focus on the moment estimation problem for multiple populations, the proposed BMF method can be generally applied to other cases where mean and variance are accurately estimated by combining the prior information from an 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 [14]. As a result, the performance distribution is fully specified by its mean μ and variance σ^2:

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

In this sub-section, we take the problem of mean estimation as an example to illustrate the BMF method proposed in [14]. Traditionally, the mean value μ in (1) is often estimated by the following estimator based on sample mean:

$$\tilde{\mu} = \frac{1}{N} \cdot \sum_{n=1}^{N} x^{(n)}, \tag{2}$$

where $\{x^{(n)}; n = 1, 2, \ldots, N\}$ denotes a set of random samples, and N is the total number of these samples.

The sample mean $\tilde{\mu}$ in (2) may not accurately approximate the actual mean μ in (1), if the sample size N is extremely small. BMF attempts to exploit the correlation between the early-stage and late-stage data to further improve the estimation accuracy. Generally, we assume that the unknown mean value μ follows a *prior distribution* of the form:

$$\mu \sim p(\mu). \tag{3}$$

On the other hand, the *likelihood function* can be expressed as:

$$p(\mathbf{x} \mid \mu), \tag{4}$$

where $\mathbf{x} = [x^{(1)}\ x^{(2)}\ \ldots\ x^{(N)}]^T$ denotes the observed samples. The likelihood function $p(\mathbf{x} \mid \mu)$ measures the probability of observing the samples $\mathbf{x}$ with a given value of μ.

Given the prior distribution $p(\mu)$ in (3) and the likelihood function $p(\mathbf{x} \mid \mu)$ in (4), *maximum-a-posteriori* (MAP) estimation can be applied to search for the optimal mean value μ by maximizing the *posterior distribution*:

$$\max_{\mu} \quad p(\mu \mid \mathbf{x}) \propto p(\mu) \cdot p(\mathbf{x} \mid \mu). \tag{5}$$

The aforementioned MAP estimation aims to find the mean value μ that is most likely to occur for the prior distribution $p(\mu)$ and the observed data $\mathbf{x}$.

2.2 Bayesian Model Fusion for Distribution Estimation

BMF has also been extended to estimate the probability density function (PDF) of a performance of interest (say, g) [12]. Due to process variations, the performance g can be modeled as a random variable that is described by its PDF $p(g)$. Without loss of generality, the function $p(g)$ can be approximated as a linear combination of a set of basis functions (e.g., trigonometric functions, orthogonal polynomials, etc.):

$$p(g) = \sum_{k=1}^{K} \alpha_k \cdot b_k(g), \tag{6}$$

where $\{b_k(g); k = 1, 2, ..., K\}$ denotes a set of basis functions, $\{\alpha_k; k = 1, 2, ..., K\}$ stands for the unknown model coefficients that should be solved, and K is the total number of basis functions.

The key idea of BMF is to borrow the early-stage knowledge to accurately estimate the late-stage PDF $p(g)$ in (6). It assumes that the late-stage PDF is similar, but not identical, to the early-stage PDF. Based on the early-stage PDF, the model coefficients $\{\alpha_k; k = 1, 2, ..., K\}$ associated with the late-stage PDF $p(g)$ are statistically characterized by a prior distribution:

$$\boldsymbol{\alpha} \sim p(\boldsymbol{\alpha}), \tag{7}$$

where $\boldsymbol{\alpha} = [\alpha_1\ \alpha_2\ \ldots\ \alpha_K]^T$ denotes the model coefficients. Next, we consider the likelihood function:

$$p(\mathbf{g} \mid \boldsymbol{\alpha}), \tag{8}$$

which represents the probability of observing a set of random samples $\mathbf{g}$ with a given value of $\boldsymbol{\alpha}$. BMF optimally estimates the unknown coefficient vector $\boldsymbol{\alpha}$ by MAP:

$$\max_{\boldsymbol{\alpha}} \quad p(\boldsymbol{\alpha} \mid \mathbf{g}) \propto p(\boldsymbol{\alpha}) \cdot p(\mathbf{g} \mid \boldsymbol{\alpha}). \tag{9}$$

After the coefficient vector $\boldsymbol{\alpha}$ is solved, the late-stage PDF of g can be easily determined by (6).

The aforementioned BMF methods for both moment estimation and distribution estimation can be efficiently applied to continuously distributed performances only, as shown in (1) and (6). If pre-silicon simulation and/or post-silicon measurement yield a binary outcome (i.e., either “pass” or “fail”), the corresponding distribution is not Gaussian or even continuous. Hence, the conventional BMF method is no longer applicable. It, in turn, motivates us to develop a new BMF method based on Bernoulli distribution (BMF-BD) to handle these special cases.

3. PROPOSED APPROACH

In this section, we develop our proposed BMF-BD method and highlight its novelties. We will first briefly describe the basics of Bernoulli distribution, and then derive the corresponding Bayesian inference for yield estimation.

3.1 Bernoulli Distribution

In general, if the outcome of pre-silicon simulation and/or post-silicon measurement is binary, it can be modeled as a Bernoulli random variable:

$$x = \begin{cases} 1 & If \quad Pass \\ 0 & If \quad Fail \end{cases}. \tag{10}$$

The statistics of x can be fully specified by its probability mass function (PMF) [18]:

$$p(x) = \begin{cases} \beta & if \quad x = 1 \\ 1-\beta & if \quad x = 0 \end{cases}, \tag{11}$$

where $\beta \in [0, 1]$ denotes the yield value that we want to estimate.

The conventional approach to determine the yield β is based on *maximum likelihood estimation* (MLE) [17]. Given a set of random samples $\mathbf{x} = [x^{(1)}\ x^{(2)}\ \ldots\ x^{(N)}]^T$ where N is the total number of these samples, MLE aims to find the optimal value of β so that the probability of observing these samples is maximized.

Towards this goal, we first re-write the PMF $p(x)$ in an

alternative form:

$$p(x \mid \beta) = \beta^{x} \cdot (1-\beta)^{1-x}, \tag{12}$$

where $x \in \{0, 1\}$. Assuming that all random samples $\{x^{(n)}; n = 1, 2, \ldots, N\}$ are generated independently, the likelihood of observing $\mathbf{x} = [x^{(1)}\ x^{(2)}\ \ldots\ x^{(N)}]^T$ is equal to:

$$p(\mathbf{x} \mid \beta) = \prod_{n=1}^{N} p\left(x^{(n)} \mid \beta\right). \tag{13}$$

Substitute (12) into (13) and we have:

$$p(\mathbf{x} \mid \beta) = \prod_{n=1}^{N} \beta^{x^{(n)}} (1-\beta)^{1-x^{(n)}}. \tag{14}$$

To find the optimal value of β that maximizes the likelihood function $p(\mathbf{x} \mid \beta)$ in (14), we further take the logarithm of (14), yielding:

$$\log\left[p(\mathbf{x} \mid \beta)\right] = \sum_{n=1}^{N}\left[x^{(n)} \cdot \log\beta + \left(1-x^{(n)}\right) \cdot \log(1-\beta)\right]. \tag{15}$$

Note that a logarithmic function is monotonically increasing. Hence, maximizing the log-likelihood function $\log[p(\mathbf{x} \mid \beta)]$ in (15) is equivalent to maximizing the original likelihood function $p(\mathbf{x} \mid \beta)$ in (14). According to the first-order optimality condition [19], the gradient of $\log[p(\mathbf{x} \mid \beta)]$ must equal zero at the optimal β:

$$\frac{\partial}{\partial \mathbf{x}} \log\left[p(\mathbf{x} \mid \beta)\right] = \mathbf{0}. \tag{16}$$

Based on (16), it is straightforward to determine the MLE value of β:

$$\beta_{MLE} = \frac{1}{N} \cdot \sum_{n=1}^{N} x^{(n)} = \frac{M}{N}, \tag{17}$$

where

$$M = \sum_{n=1}^{N} x^{(n)} \tag{18}$$

is the total number of samples that pass the specification.

The aforementioned MLE method results in an unbiased estimator for β without relying on any prior knowledge. In practice, however, since we can learn prior information from the early stage, MAP often offers superior accuracy over MLE, especially if the statistics at the late stage must be estimated with very few random samples [12]-[14]. For instance, consider a circuit example for which the yield is 99% (i.e., $\beta = 0.99$). In this case, MLE is likely to require hundreds of random samples in order to observe a sufficient number of "failed" samples to accurately estimate the yield. On the other hand, if prior knowledge can be appropriately learned from the early stage, MAP can accurately estimate the yield with less than 100 samples, as will be demonstrated by our circuit examples in Section 4. In what follows, we will derive the detailed formulation of our proposed BMF-BD method to estimate yield by MAP for Bernoulli distribution.

3.2 Prior Knowledge Definition

BMF-BD assumes that the yield values at early and late stages are similar but not identical. In order to encode our prior knowledge that can be learned from the early stage, we must appropriately define a prior distribution. However, unlike other conventional BMF methods where the prior distribution and the likelihood function are both modeled as Gaussian distributions [12]-[14], such a simple Gaussian prior is not applicable here, because our likelihood function for yield estimation is a Bernoulli distribution, as shown in (14). If a Gaussian prior is used in this case, the posterior distribution does not have a closed-form expression and, hence, MAP cannot be easily applied to solve the Bayesian interference [17].

To address this technical challenge, we borrow the theory of conjugate prior from the statistics community [17]. If a prior distribution and a posterior distribution are in the same family (e.g., both are Gaussian distributions), they are called conjugate distributions and the prior distribution is called a conjugate prior for the likelihood function. It has been shown by the statistics community that the conjugate prior of Bernoulli distribution is a Beta distribution [17]:

$$p(\beta \mid a, b) = \frac{\Gamma(a+b)}{\Gamma(a) \cdot \Gamma(b)} \cdot \beta^{a-1} \cdot (1-\beta)^{b-1} \quad (0 \le \beta \le 1), \tag{19}$$

where $\Gamma(\bullet)$ denotes the gamma function, and $a \ge 1$ and $b \ge 1$ are two hyper-parameters that control the shape of the distribution. Figure 1 shows several different Beta distributions with different values of a and b.

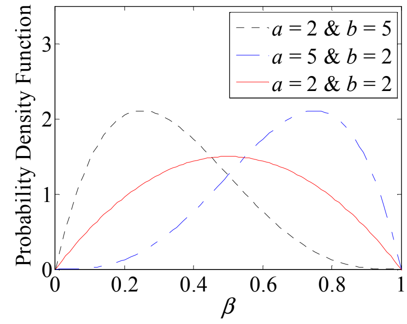

Figure 1. Several different Beta distributions are shown with different values of a and b.

Studying Figure 1 reveals three important properties of a Beta distribution. First, a Beta distribution $p(\beta \mid a, b)$ is defined over the interval $\beta \in [0, 1]$ that covers all possible yield values. Second, $p(\beta \mid a, b)$ is peaked at a particular value (say, β_M) that is defined as the mode of the distribution [17]. It can be proven that β_M is a simple function of a and b [17]:

$$\beta_M = \frac{a-1}{a+b-2}. \tag{20}$$

Third, as β moves away from β_M, the PDF value decays quickly, meaning that it is unlikely to observe a β value that is substantially different from β_M.

Based on the aforementioned discussions, we propose to use the Beta distribution $p(\beta \mid a, b)$ in (19) to encode our prior knowledge to estimate the late-stage yield β. In addition, we further set up the following constraint for the hyper-parameters a and b:

$$\beta_M = \frac{a-1}{a+b-2} = \beta_E, \tag{21}$$

where β_E denotes the yield value at the early stage. The prior distribution $p(\beta \mid a, b)$ defined by (19) and (21) implies that the late-stage yield β is likely to be similar to β_E, since $p(\beta \mid a, b)$ is peaked at $\beta_M = \beta_E$. In other words, it is unlikely for β and β_E to be extremely different and the beta distribution $p(\beta \mid a, b)$ appropriately encodes our prior knowledge that the early-stage and late-stage data are strongly correlated.

Eq. (21) can be re-written as:

$$b = \frac{(1-\beta_E) \cdot (a-1)}{\beta_E} + 1. \tag{22}$$

Substituting (22) into (19) yields:

$$p(\beta \mid a)=\frac{\Gamma\left((a-1)/\beta_E+2\right)\cdot\beta^{a-1}\cdot(1-\beta)^{(1-\beta_E)\cdot(a-1)/\beta_E}}{\Gamma(a)\cdot\Gamma\left((1-\beta_E)\cdot(a-1)/\beta_E+1\right)}. \quad (23)$$

In (23), there is only one hyper-parameter a that should be determined. The optimal value of a can be estimated by an MLE method, as will be discussed in detail in Section 3.4.

3.3 Maximum-A-Posteriori Estimation

After defining the prior distribution $p(\beta \mid a)$ in (23), we combine (23) with few late-stage samples to estimate the late-stage yield β by MAP. MAP further reduces the uncertainty of the prior distribution by taking into account the samples that become available at the late stage.

Given a set of late-stage samples $\{x^{(n)}; n = 1, 2, \ldots, N\}$, the vector $\mathbf{x}$ follows a Bernoulli distribution, as shown in (14). Based on Bayes' theorem, the posterior distribution $p(\beta \mid \mathbf{x})$ is proportional to the prior distribution $p(\beta \mid a)$ in (23) multiplied by the likelihood function $p(\mathbf{x} \mid \beta)$ in (14):

$$p(\beta \mid \mathbf{x}) \propto \beta^{M+a-1}(1-\beta)^{N-M+(1-\beta_E)\cdot(a-1)/\beta_E}, \quad (24)$$

where the value of M is defined in (18). After appropriately normalizing the right-hand side of (24), we can write the posterior distribution $p(\beta \mid \mathbf{x})$ as:

$$p(\beta|\mathbf{x})=\frac{\Gamma\left(N+(a-1)/\beta_E+2\right)\beta^{M+a-1}(1-\beta)^{N-M+(1-\beta_E)(a-1)/\beta_E}}{\Gamma(M+a)\Gamma\left(N-M+(1-\beta_E)(a-1)/\beta_E+1\right)}. \quad (25)$$

The posterior distribution $p(\beta \mid \mathbf{x})$ in (25) models the uncertainty of β after observing the late-stage data $\{x^{(n)}; n = 1, 2, \ldots, N\}$. Note that $p(\beta \mid \mathbf{x})$ in (25) remains a Beta distribution. The key idea of MAP is to find the value of β_{MAP} at which the posterior distribution $p(\beta \mid \mathbf{x})$ reaches its maximum. Namely, we want to find the optimal value β_{MAP} that is most likely to occur based on the posterior probability. Hence, β_{MAP} is simply the mode of the Beta distribution $p(\beta \mid \mathbf{x})$ in (25):

$$\beta_{MAP}=\frac{M+a-1}{N+(a-1)/\beta_E}. \quad (26)$$

Studying (26) reveals an important fact that the hyper-parameter a plays an important role in determining β_{MAP}. In one extreme case, if a is sufficiently large, we have:

$$\beta_{MAP} \approx \beta_E. \quad (27)$$

In this case, β_{MAP} is approximately equal to the early-stage yield β_E, implying that the prior knowledge is sufficiently accurate and the late-stage samples are almost unused to determine the late-stage yield. In the other extreme case, if a is sufficiently small, we have:

$$\beta_{MAP} \approx \frac{M}{N}. \quad (28)$$

In this case, β_{MAP} is approximately equal to the MLE estimation β_{MLE} in (17), meaning that the prior knowledge is not correct and is almost neglected by the Bayesian inference.

The aforementioned discussion demonstrates a strong need to appropriately set the value of a before performing the MAP estimation in (26). In the next sub-section, we will present a novel statistical method to fulfill this goal.

3.4 Hyper-parameter Estimation

The hyper-parameter a is an important variable in our Bayesian inference. It controls the shape of the prior distribution and, hence, the confidence of our prior knowledge. For instance, if a is large, the prior distribution $p(\beta \mid a)$ in (23) is narrowly peaked around its mode β_E. In this case, the prior knowledge is expected to be accurate and, hence, the MAP estimation β_{MAP} is approximately equal to β_E, as shown in (27).

In practice, however, we do not know the "accuracy" of the prior knowledge that is expected to be case-dependent. Namely, the optimal value of a may substantially vary from case to case. It is impossible to set a to a constant value in advance. Instead, it must be adaptively "learned" from the given data at the late stage.

To derive our statistical method for solving the hyper-parameter a, we combine the prior distribution $p(\beta \mid a)$ in (23) and the likelihood function $p(\mathbf{x} \mid \beta)$ in (14), resulting in the joint distribution:

$$\begin{aligned} p(\mathbf{x},\beta \mid a) &= p(\mathbf{x} \mid \beta)\cdot p(\beta \mid a) \\ &= \frac{\Gamma\left((a-1)/\beta_E+2\right)\beta^{M+a-1}(1-\beta)^{N-M+(1-\beta_E)(a-1)/\beta_E}}{\Gamma(a)\Gamma\left((1-\beta_E)(a-1)/\beta_E+1\right)}, \end{aligned} \quad (29)$$

where M is a function of the late-stage data $\{x^{(n)}; n = 1, 2, \ldots, N\}$, as shown in (18). Since β in (29) represents the yield, its value must be within the interval $\beta \in [0, 1]$. Integrating the joint distribution $p(\mathbf{x}, \beta \mid \alpha)$ in (29) over β yields the marginal distribution:

$$\begin{aligned} p(\mathbf{x}|a) &= \int_0^1 p(\mathbf{x},\beta|a)\cdot d\beta \\ &= \frac{\Gamma\left((a-1)/\beta_E+2\right)\Gamma(M+a)\Gamma\left(N-M+(1-\beta_E)(a-1)/\beta_E+1\right)}{\Gamma(a)\Gamma\left((1-\beta_E)(a-1)/\beta_E+1\right)\Gamma\left(N+(a-1)/\beta_E+2\right)}. \end{aligned} \quad (30)$$

Eq (30) is an important equation, since it tells us the probability of observing the late-stage data $\mathbf{x}$ for a given value of a. Based on (30), we propose to adopt a maximum likelihood strategy to estimate the value of a. Namely, we aim to find the optimal value of a that maximizes the likelihood function $p(\mathbf{x} \mid \alpha)$ in (30):

$$\max_a \frac{\Gamma\left((a-1)/\beta_E+2\right)\Gamma(M+a)\Gamma\left(N-M+(1-\beta_E)(a-1)/\beta_E+1\right)}{\Gamma(a)\Gamma\left((1-\beta_E)(a-1)/\beta_E+1\right)\Gamma\left(N+(a-1)/\beta_E+2\right)}. \quad (31)$$

Even though the optimization problem in (31) may not be convex, it is one-dimensional and, hence, can be easily solved by a linear search algorithm [20]. In other words, we can evaluate the merit function in (31) for a list of possible values of a, and then select the optimal value that maximizes the merit function.

Appropriately determining the optimal value of a by solving (31) allows us to quantitatively assess the accuracy of our prior knowledge. Most importantly, if the prior knowledge is not accurate, it should be ignored (i.e., by setting a small value for a) when estimating the late-stage yield. As such, the proposed BMF-BD method is not biased by the "wrong" prior information and, hence, can be robustly applied to a broad range of practical applications, as will be demonstrated by the experimental results in Section 4.

3.5 Summary

Algorithm 1: BMF-BD for Yield Estimation

1. Start from a given early-stage yield β_E and a set of late-stage samples $\{x^{(n)}; n = 1, 2, \ldots, N\}$.
2. Calculate the value of M by using (18).
3. Apply linear search to solve the optimization problem in (31) to find the optimal value of a based on MLE.
4. Determine the late-stage yield β_{MAP} in (26) based on MAP.

Algorithm 1 summarizes the major steps of our proposed BMF-BD algorithm for yield estimation. Starting from a given early-stage yield β_E as our prior knowledge and a set of late-stage samples $\{x^{(n)}; n = 1, 2, \ldots, N\}$, we first count the number of late-

stage samples that pass the specification, i.e., the value of M defined in (18). Next, we estimate the hyper-parameter a for our Beta prior $p(\beta \mid a)$ in (23), where the optimal value of a is determined by solving the one-dimensional optimization problem in (31) based on MLE. Once the prior distribution $p(\beta \mid a)$ in (23) is fully specified with the optimal a value, we combine the prior distribution $p(\beta \mid a)$ with the late-stage samples $\{x^{(n)}; n = 1, 2, \ldots, N\}$ to estimate the late-stage yield β_{MAP} in (26) based on MAP. As will be demonstrated by our circuit examples in Section 4, the proposed BMF-BD method achieves up to 10× cost reduction over the conventional approach without surrendering any accuracy.

4. NUMERICAL EXAMPLES

In this section, two circuit examples are used to demonstrate the efficiency of our proposed BMF-BD method. For testing and comparison purposes, two different yield estimation methods are implemented: (i) the conventional maximum likelihood estimation (MLE), and (ii) the proposed Bayesian model fusion (BMF-BD). All experiments are performed on a server with 2.66GHz dual-core CPU and 4GB memory.

4.1 SRAM Read Path

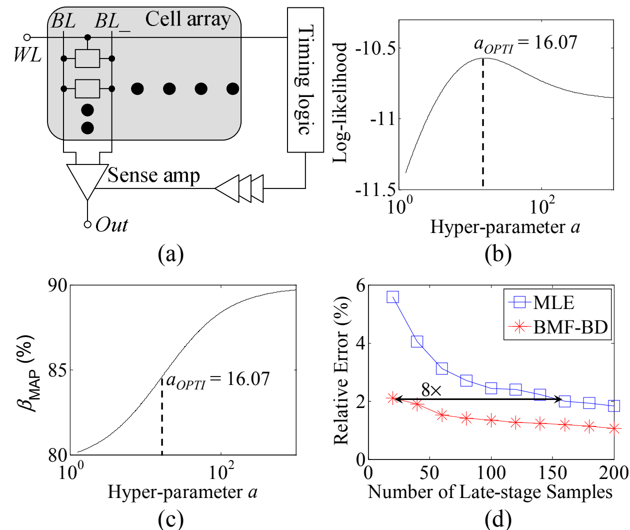

Figure 2. (a) A simplified schematic is shown for an SRAM read path designed in a 65nm CMOS process. (b) The log-likelihood $\log[p(\mathbf{x} \mid \alpha)]$ is plotted as a function of the hyper-parameter a with 20 late-stage samples where the likelihood is maximized at a_{OPTI} = 16.07. (c) The estimated late-stage yield β_{MAP} is plotted as a function of the hyper-parameter a with 20 late-stage samples where the optimal value of β_{MAP} is 84.51% at a_{OPTI} = 16.07. (d) The estimation error of late-stage yield is calculated from 200 repeated runs and plotted as a function of the number of late-stage samples.

Figure 2(a) shows the simplified schematic of an SRAM read path designed in a 65nm CMOS process. The read path consists of three major components: the SRAM cell array, the timing logic and the sense amplifier. Once the word-line (*WL*) is activated to turn on a particular SRAM cell, the two bit-lines *BL* and *BL_* are discharged to develop a small voltage difference. The sense amplifier is then used to compare *BL* and *BL_* and generate a binary decision at the output (i.e., either "0" or "1"). Due to process variations, the SRAM circuit may produce a wrong output and, hence, the read operation fails. In this example, we consider the schematic-level design as the early stage and the post-layout design as the late stage. Our objective is to accurately estimate the late-stage yield by borrowing the prior knowledge from the early stage.

We first collect 5000 early-stage (i.e., schematic-level) Monte Carlo samples and estimate the early-stage yield β_E = 89.88%, where the device-level variations of all transistors are considered. In practice, we assume that the early-stage data are already available when estimating the late-stage yield. Therefore, there is no additional cost to re-use the early-stage data. Next, we generate 20 late-stage (i.e., post-layout) Monte Carlo samples. The SRAM circuit produces correct outputs at 16 out of these 20 late-stage samples. Hence, the late-stage yield estimated by the conventional MLE is simply β_{MLE} = 16/20 = 80%. For testing purposes, we further collect 5000 late-stage Monte Carlo samples to calculate the "actual" late-stage yield β_{EXACT} = 90.66%.

On the other hand, we apply our proposed BMF-BD to estimate the late-stage yield from the same 20 late-stage samples. Figure 2(b) shows the log-likelihood $\log[p(\mathbf{x} \mid \alpha)]$ as a function of the hyper-parameter a. Note that $\log[p(\mathbf{x} \mid \alpha)]$ is peaked at a = 16.07. As the value of a moves away from 16.07, the likelihood function decreases. Hence, the optimal value of the hyper-parameter is a_{OPTI} = 16.07 in this case.

Figure 2(c) plots the late-stage yield β_{MAP} estimated by BMF-BD as a function of the hyper-parameter a. Studying Figure 2(b) reveals two important observations. First, as the value of a increases from a small value (e.g., 1) to a large value (e.g., 1000), the value of β_{MAP} varies from β_{MLE} (i.e., the MLE estimation) to β_E (i.e., the early-stage yield). This observation is consistent with our intuition discussed in Section 3.3. Second, since the optimal hyper-parameter a_{OPTI} equals 16.07 in this case, the late-stage yield estimated by BMF-BD is β_{MAP} = 84.51%. Even though β_{MAP} is not exactly identical to β_{EXACT} = 90.66%, it is substantially more accurate than β_{MLE} = 80% in this example.

To further compare the accuracy between MLE and BMF-BD, we vary the number of late-stage (i.e., post-layout) samples from 20 to 200 and calculate the estimation error of late-stage yield:

$$Error = \left| \frac{\beta_{ESTI} - \beta_{EXACT}}{\beta_{EXACT}} \right|, \tag{32}$$

where β_{ESTI} denotes the estimated yield (i.e., β_{MLE} by MLE or β_{MAP} by BMF-BD). Figure 2(d) plots the error as a function of the number of late-stage samples. To average out random fluctuations, the error in Figure 2(d) is calculated from 200 repeated runs based on independent samples. Note that BMF-BD is able to offer high accuracy for yield estimation, even if the number of late-stage samples is as low as 20. To achieve the same accuracy, MLE requires 160 late-stage samples. Since the computational time of post-layout simulation dominates the yield estimation cost, BMF-BD achieves 8× runtime speed-up over MLE in this example.

4.2 Silicon Measurement Data

In this example, we consider the silicon measurement data collected by a major semiconductor company from two different tape-outs. The total numbers of silicon dies measured from these two tape-outs are 2305 and 2010, respectively. Each die is tested and labeled as either "pass" or "fail". The early-stage (i.e., from the first tape-out) and late-stage (i.e., from the second tape-out) yield values are β_E = 90.63% and β_{EXACT} = 90.25%, respectively. Our objective is to take the early-stage yield β_E as the prior

knowledge to accurately estimate the late-stage yield with very few late-stage measurements.

Figure 3(a) plots the estimation error of late-stage yield as a function of the number of late-stage samples. Similar to the previous example, BMF-BD offers superior accuracy for yield estimation over MLE. To achieve the same accuracy, BMF-BD only requires 20 late-stage samples, while MLE requires 200 samples (10× more). In this example, BMF-BD is extremely accurate, since the early-stage yield β_E is almost identical to the late-stage yield β_{EXACT} and, hence, provides accurate prior knowledge for late-stage yield estimation.

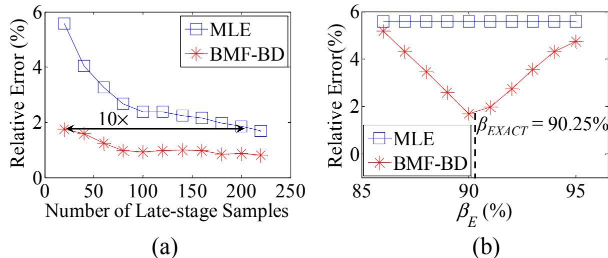

Figure 3. The estimation error of late-stage yield is calculated from 200 repeated runs and plotted as a function of (a) the number of late-stage samples, and (b) the early-stage yield β_E defining our prior knowledge.

For testing purposes, we set the early-stage yield β_E to different values and Figure 3(b) shows the resulting estimation error of late-stage yield as a function of β_E. Note that the error increases, as β_E deviates from the actual late-stage yield β_{EXACT}. However, even if the absolute difference between β_E and β_{EXACT} reaches 5% (i.e., our prior knowledge is not highly accurate), BMF-BD still out-performs MLE. It, in turn, demonstrates the robustness of the proposed BMF-BD method for practical applications.

5. CONCLUSIONS

In this paper, a novel method of Bayesian model fusion on Bernoulli distribution (BMF-BD) is proposed for efficient yield estimation where the pre-silicon simulation and/or post-silicon measurement results are binary. BMF-BD models the binary simulation/measurement outcome as a Bernoulli distribution. In addition, it further encodes the prior knowledge from an early stage as a Beta distribution which is the conjugate prior of Bernoulli distribution. By combining the prior knowledge with very few late-stage data through Bayesian inference, BMF-BD can predict the late-stage yield with extremely high accuracy. Our circuit examples demonstrate that BMF-BD achieves up to 10× cost reduction over the conventional estimator without surrendering any accuracy. BMF-BD can be further incorporated into a yield monitoring flow for design optimization and/or process tuning.

6. ACKNOWLEDGEMENTS

This research is supported partly by the National Natural Science Foundation of China (NSFC) research project 91330201, 61125401, 61376040 and 61228401, partly by the National Basic Research Program of China under the grant 2011CB309701, partly by the National Major Science and Technology Special Project 2014ZX02301001-005-002 of China during the 12-th five-year plan period, partly by Shanghai Science and Technology Committee project 13XD1401100, partly by "Chen Guang" project supported by Shanghai Municipal Education Commission and Shanghai Education Development Foundation.

7. REFERENCES

[1] X. Li, et al., *Statistical Performance Modeling and Optimization*, Now Publishers, 2007.

[2] Semiconductor Industry Associate, *International Technology Roadmap for Semiconductors*, 2011.

[3] A. Srivastava, et al., "Accurate and efficient gate-level parametric yield estimation considering correlated variations in leakage power and performance," *IEEE DAC*, pp. 535-540.

[4] X. Li, et al., "Asymptotic probability extraction for nonnormal performance distributions," *IEEE Trans. on CAD*, vol. 26, no. 1, pp. 16-37, Jan. 2007.

[5] 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.

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

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

[8] J. Rivers, et al., "Error tolerance in server class processors," *IEEE Trans. on CAD*, vol. 30, no. 7, pp. 945-959, Jul. 2011.

[9] A. Tang, et al., "A low-overhead self-healing embedded system for ensuring high yield and long-term sustainability of 60GHz 4Gb/s radio-on-a-chip", *IEEE ISSCC*, pp. 316-318, 2012.

[10] B. Sadhu, et al., "A linearized, low-phase-noise VCO-based 25GHz PLL with autonomic biasing," *IEEE JSSC*, vol. 48, no. 5, pp. 1138-1150, May. 2013.

[11] P. Gupta, et al., "Underdesigned and opportunistic computing in presence of hardware variability," *IEEE Trans. on CAD*, vol. 32, no. 1, pp. 8-23, Jan. 2013.

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

[13] 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.

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

[15] S. Sun, et al., "Indirect performance sensing for on-chip analog self-healing via Bayesian model fusion," *IEEE CICC*, 2013.

[16] X. Li, et al., "Bayesian model fusion: a statistical framework for efficient pre-silicon validation and post-silicon tuning of complex analog and mixed-signal circuits," *IEEE ICCAD*, pp. 795-802, 2013.

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

[18] A. Papoulis, et al., *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, 2001.

[19] S. Boyd, et al., *Convex Optimization*. Cambridge University Press, 2004.

[20] D. Knuth, *The Art of Computer Programming: Sorting and Searching*, Addison-Wesley, 1998.