

# Efficient Bit Error Rate Estimation for High-Speed Link by Bayesian Model Fusion

Chenlei Fang<sup>1</sup>, Qicheng Huang<sup>1</sup>, Fan Yang<sup>1,\*</sup>, Xuan Zeng<sup>1,\*</sup>, Xin Li<sup>1,2</sup> and Chenjie Gu<sup>3</sup>

<sup>1</sup>State Key Lab of ASIC & System, Microelectronics Department, Fudan University, Shanghai, P. R. China

<sup>2</sup>Electrical & Computer Engineering Department, Carnegie Mellon University, Pittsburgh, PA, USA

<sup>3</sup>Intel Strategic CAD Labs

**Abstract**—High-speed I/O link is an important component in computer systems, and estimating its bit error rate (BER) is a critical task to guarantee its performance. In this paper, we propose an efficient method to estimate BER by Bayesian Model Fusion. Its key idea is to borrow conventional extrapolated BER value as prior knowledge, and combine it with additional measurement data to “calibrate” the BER value. This method can be viewed as an application of Bayesian Model Fusion (BMF) technique. We further propose some novel methodologies to make BMF applicable in the BER estimation case. In this way, we can sufficiently decrease the number of bits needed to estimate BER value. Several experiments demonstrate that our proposed method achieves up to 8x speed-up over direct estimation method.

## I. INTRODUCTION

High-speed I/O link is a critical component in modern computer systems. It connects different parts of a digital system (e.g., between CPU and DRAM), and enables high-speed data communication among these components. Efficiently designing and testing a high-speed link to guarantee sufficiently small bit error rate (BER), however, is an extremely challenging task. In particular, since the BER of a high-speed link is often in the order of  $10^{-12}$ , it requires us to transmit a large number of (e.g.,  $10^{13}$ ) bits in order to accurately estimate such a small BER value, thereby resulting in expensive testing cost. For this reason, an efficient-yet-accurate methodology must be used for BER estimation with low measurement cost.

Towards this goal, a variety of techniques have been proposed in the literature. Among them, BER extrapolation is one of the promising methods [2-5]. It measures a set of relatively large (e.g., greater than  $10^{-10}$ ) BER values with different time and/or voltage margins. Next, the actual BER is estimated by extrapolation based on an analytical model. While the conventional extrapolation method has been successfully applied to many practical applications, the underlying analytical model is derived with various simplifications. Hence, the estimated BER value by extrapolation may substantially deviate from the actual one.

To address this issue, we propose a novel statistical method to further correct the estimation error introduced by BER extrapolation. Our key idea is to first consider the extrapolated BER value as our prior knowledge. Next, we further transmit a number of bits without adding any time and/or voltage margin, and measure the output of the receiver. The additional measurement data is finally combined with the prior knowledge to “calibrate” the BER value via Bayesian inference. From this point of view, our proposed BER estimation attempts to reduce the error of BER extrapolation

by taking advantage of the additional measurement data we collect.

The aforementioned BER estimation for high-speed link can be viewed as a new application of the Bayesian model fusion (BMF) technique that has been recently developed in the literature [6-10]. In addition, a number of novel algorithms and methodologies are proposed in this paper in order to make BMF applicable to our problem of BER estimation. First, a statistically optimal method based on maximum likelihood estimation (MLE) is developed to estimate the BER by extrapolation. The MLE approach can statistically model the “confidence” of the measured BER values with different time and/or voltage margins and appropriately weight them for BER extrapolation.

Second, since the BER of interest is extremely small (e.g.,  $10^{-12}$ ), directly applying the BMF algorithm in [10] suffers from numerical issues. To improve numerical stability, a new approximation scheme is proposed in this paper to compute the Bayesian inference for BER calibration and, therefore, make our proposed method applicable to practical applications.

The remainder of this paper is organized as follows. Section II briefly summarizes the background for BER extrapolation and BMF, and then our proposed method is described in Section III. Several implementation details are further discussed in Section IV. The efficacy of the proposed method is demonstrated by an industrial high-speed link example with silicon measurement data in Section V. Finally, we conclude in Section VI.

## II. BACKGROUND

### A. BER Extrapolation

The bit error rate or bit error ratio (BER) is defined as the number of bit errors divided by the total number of transferred bits during a studied time interval. It is not practical to directly estimate extremely small BER, because it would take a long time to observe a bit error. BER extrapolation method has been proposed to address this problem [2-5], which extrapolate the BER from a few points that require less time to measure. The extrapolation method is usually implemented on a Bit Error Rate Tester (BERT).

Traditionally, in order to estimate the BER, the data is sampled at the optimal point of the eye diagram, where the eye opening is largest. The threshold of logic “0” and “1” is also set to the middle of the amplitude. The corresponding BER value is usually very small (e.g.  $10^{-12}$ ).

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

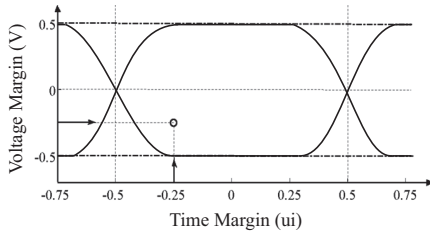


Fig. 1. Sample point with both time and voltage margin

To reduce the time for BER measurement, we can artificially increase the BER by sampling the data with different configurations. For example, the data can be sampled at a time point with offsets to the ideal sampling point. The offsets are also called time margins. Alternatively, the threshold for logic “0” and “1” can be set to a voltage with offsets to the middle of the amplitude. Such an offsets are also called voltage margins. Fig. 1 shows an example of time (x-axis) and voltage (y-axis) margins.

By measuring the BER with a few different time/voltage margins, we can extrapolate the accurate BER from these artificial BERs. We take the BER extrapolation from BERs with time margins as an example here to illustrate the BER extrapolation method. The BER extrapolation from BERs with voltage margins can be derived similarly.

If we know the jitter distribution, the BER values with different time margins can be predicted. A commonly used approximation for jitter distribution is the dual-Dirac model. The dual-Dirac model assume that the total jitter is a convolution of the random jitter (RJ) with Gaussian distribution and the deterministic jitter (DJ) with dual-Dirac distribution, as demonstrated in Fig. 2. In Fig. 2,  $\mu_L$  and  $\mu_R$  are the left and right components of the dual-Dirac distribution, respectively. The jitter distribution can be divided into three regions: (1) at the crossing-point the distribution is dominated by DJ; (2) at time-delays farther from the crossing-point the distribution is increasingly dominated by RJ; (3) far from the crossing point, in the asymptotic limit, the tails follow the Gaussian RJ distribution. The bit errors occur at the two tails of the distribution far from the crossing point, which follow the Gaussian RJ distribution.

The BER with time offset  $m$  can thus be expressed as an integral of the Gaussian RJ distribution from  $m$  to infinity

$$BER(m) = \frac{1}{\sqrt{2\pi}\sigma} \int_m^{\infty} \exp\left[-\frac{(\mu_L - x)^2}{2\sigma^2}\right] dx. \quad (1)$$

According to (1), we can find that BER is a nonlinear function of the time margin  $m$ , which makes the extrapolation very complicated. To address this problem, a nonlinear transformation can be introduced to map the  $BER(m)$  to a new function  $Q(m)$  in (2):

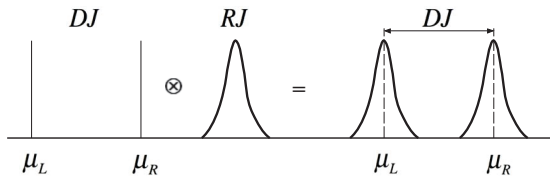


Fig. 2. Total jitter distribution in dual-Dirac model

$$Q(m) = \sqrt{2} \operatorname{erfc}^{-1}(2BER(m)). \quad (2)$$

where  $\operatorname{erfc}(\bullet)$  denotes the complementary error function. It can be proved that  $Q$  is linear to the time margin  $m$

$$Q = \frac{\mu_L - m}{\sigma}. \quad (3)$$

Once the BER values with different time margins are measured, they are transformed to  $Q$ -domain. Least-square method can be used to extrapolate  $Q(0)$  with zero margin, which corresponds to the  $Q$  value of the accurate BER. With an inverse transformation of (2), we can get the accurate BER

$$BER(0) = \frac{1}{2} \operatorname{erfc}\left(\frac{Q(0)}{\sqrt{2}}\right). \quad (4)$$

The accurate BER is thus obtained with less measurement time. This extrapolation method is widely used in industry. However, the Gaussian approximation of the tails may be inaccurate, as illustrated in Fig. 3. As a result, the BER obtained by extrapolation may be inaccurate.

### B. Bayesian Model Fusion

Bayesian Model Fusion (BMF) has been proposed to accurately estimate the statistical performances of integrated circuits, including the distribution of circuit performances and yield. BMF borrows the simulation and/or measurement data from earlier stages of design to help estimate statistical performances of later stages. Compared to traditional methods such as Monte Carlo, fewer late stage samples are needed to give an accurate estimation. The cost of verification and/or validation is thus greatly reduced.

BMF has been extended to estimate the yield where the pre-silicon simulation and/or post-silicon measurement results are binary, which is called BMF-BD method [10]. If the outcome of a simulation/measurement is binary, the result can be modeled as a random variable with Bernoulli distribution

$$x = \begin{cases} 1 & \text{if pass} \\ 0 & \text{if fail} \end{cases}. \quad (5)$$

Its probability density function can be expressed as

$$p(x|\beta) = \beta^x \cdot (1-\beta)^{1-x}, \quad (6)$$

where  $\beta \in [0, 1]$  denotes the yield to be estimated.

As early stage data and late stage data are taken from the same or similar circuit designs, BMF-BD method assumes that the early stage yield is similar, but not identical to the late stage yield. In order to correctly encode the prior knowledge from early stage yield, a prior distribution of  $\beta$  is defined as

$$p(\beta|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \beta^{a-1} \cdot (1-\beta)^{b-1} \quad (0 \leq \beta \leq 1), \quad (7)$$

which is known as Beta distribution. It is the conjugate prior of a Bernoulli distribution, which makes the prior distribution and the posterior distribution in the same family. Here,  $\Gamma(\bullet)$  denotes Gamma function, and  $a \geq 1$ ,  $b \geq 1$  are called hyper-parameters to control the shape of the Beta distribution[11].

The likelihood function of a set of observed samples  $x = [x^{(1)} x^{(2)} \dots x^{(N)}]^T$  can be expressed as

$$p(\mathbf{x} | \beta) = \prod_{n=1}^N \beta^{x^{(n)}} (1-\beta)^{1-x^{(n)}}. \quad (8)$$

BMF-BD derives the yield estimation by maximizing the posterior distribution of  $\beta$

$$\max_{\beta} p(\beta | \mathbf{x}) \propto p(\beta) \cdot p(\mathbf{x} | \beta). \quad (9)$$

The yield estimation of the binary outcome is similar to the BER testing process, as whether a received bit is “the same” or “different” from the sent bit can also be modeled as a Bernoulli random variable. As a result, BMF-BD method could be employed to improve the efficiency of BER estimation.

### III. PROPOSED APPROACH

The BER obtained by extrapolation is not accurate enough, while the direct BER measurement is not practical. We propose to take a few additional measurements data to “calibrate” the extrapolation value with BMF method.

The extrapolation BER is taken as our prior knowledge. The relationships between BER and voltage/time margins are similar. Without loss of generality, we take BER extrapolation with time margins as an example to illustrate the idea of BMF. After taking measurements at several time margins  $\{m_1, m_2, \dots, m_k\}$ , we apply the transformation (2) and obtain the  $Q$  values  $\{Q_1, Q_2, \dots, Q_k\}$  for extrapolation

$$Q_k = \sqrt{2} \operatorname{erfc}^{-1} [1 - 2P_k(m_k)], \quad (10)$$

where  $P_k$  represents the BER at the  $k$ -th margin  $m_k$  and  $Q_k$  stands for the corresponding  $Q$  value. After performing linear regression on these sample points, the regression model between  $Q$  and the time margin  $m$  can be expressed as

$$Q = cm + d, \quad (11)$$

where  $c$  and  $d$  are parameters obtained by regression. By substituting  $m = 0$  into (11) and perform the inverse transformation as (4), the BER  $\rho_e$  can be obtained. We take  $\rho_e$  as the prior knowledge for Bayesian inference.

Bit error detection can be modeled as a Bernoulli random variable:

$$x = \begin{cases} 1 & \text{if received bit is right} \\ 0 & \text{if received bit is wrong} \end{cases}. \quad (12)$$

Using this definition, BMF proposed in [10] can be applied here. Let  $\beta$  denote the probability that the received bit is the same as the sent bit. BER and  $\beta$  value follows the relationship

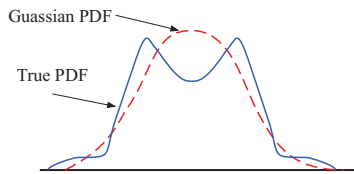


Fig. 3. Real jitter/variation distribution may not be Gaussian

$$BER = 1 - \beta. \quad (13)$$

We set the  $\beta$  value here with the same form of prior distribution as shown in (7). Given the BER by extrapolation as  $\rho_e$ , since the probability function  $p(\beta|a,b)$  peaks at a particular value defined as the mode of the distribution, we set up the following constraint for the hyper-parameters:

$$\rho_e = 1 - \beta_{MODE} = 1 - \frac{a-1}{a+b-2}, \quad (14)$$

This prior distribution implies that the real BER value is unlikely to be largely different from the extrapolation value. Substituting the constraint (14) to (7), we have the prior knowledge in this form:

$$p(\beta|a) = \frac{\Gamma((a-1)/(1-\rho_e)+2) \cdot \beta^{a-1} \cdot (1-\beta)^{(a-1)\rho_e/(1-\rho_e)}}{\Gamma(a) \cdot \Gamma((a-1)\rho_e/(1-\rho_e)+1)}, \quad (15)$$

Only one hyper-parameter  $a$  is in the final distribution, which can be determined by an MLE method described in Section IV B.

Sample measurements without time/voltage margins are further collected to calibrate the BER value. Suppose there are  $N_c$  bits are received correctly in  $N$  transmitted samples, according to Bayesian theory, the posterior distribution is proportional to the early distribution in (15) multiplying the likelihood function in (8):

$$p(\beta | \mathbf{x}) \propto \beta^{N_c+a-1} \cdot (1-\beta)^{N-N_c+(a-1)\rho_e/(1-\rho_e)}. \quad (16)$$

It can be proved that, after normalization, the posterior distribution remains a Beta distribution. Maximum-a-posterior (MAP) method estimates the  $\beta$  value or the “calibrated” BER value by finding the value where the posterior distribution reaches its maximum, hence, the mode value:

$$BER_{MAP} = 1 - \frac{N_c + a - 1}{N + (a-1)/(1-\rho_e)}, \quad (17)$$

Similar to the yield case in [10], (17) demonstrates that the “calibrated” BER value is made up of two parts: the prior knowledge given by extrapolation, and real observations. Our proposed method combines advantages of both methods: it gives remarkably accurate estimation with fewer observed sample points.

### IV. IMPLEMENTATION DETAILS

We propose several novel techniques to accelerate the BMF process for BER estimation. These techniques will be discussed in detail in this section.

#### A. Maximum Likelihood Estimation

The idea of extrapolation has been reviewed in Section II. Traditionally, the extrapolation model in the form of (11) is obtained by regression. In this paper, we develop a statistically optimal algorithm to get the linear coefficients  $c$  and  $d$  by maximum likelihood estimation (MLE).

Bit error detection can be modeled as a Bernoulli random variable as shown in (12). Assume that  $N_k$  random samples are

collected for the  $k$ -th margin value  $m_k$ , the BER value  $P_k^{MC}$  with margin  $m_k$  can be expressed as

$$P_k^{MC} = \frac{1}{N_k} \cdot \sum_{n=1}^{N_k} (1 - x_n). \quad (18)$$

The variance of the estimator  $P_k^{MC}$  can be written as

$$\sigma_k^2 = \frac{1}{N_k} \cdot P_k^{MC} (1 - P_k^{MC}). \quad (19)$$

The estimator  $P_k^{MC}$  follows a Normal distribution according to the central limit theorem

$$P_k^{MC} \sim Gauss(P_k, \sigma_k^2), \quad (20)$$

where  $P_k$  is the mean of the estimator  $P_k^{MC}$ .

$Q_k^{MC}$  also follows a Normal distribution

$$Q_k^{MC} \sim Gauss(Q_k, \sigma_{Q,k}^2), \quad (21)$$

where  $Q_k$  and  $\sigma_{Q,k}^2$  denote the mean and variance of  $Q_k^{MC}$ , respectively.

Taking the first-order Taylor expansion of (10) at  $P_k$ , the estimator  $Q_k^{MC}$  can be written as

$$\begin{aligned} Q_k^{MC} &\approx Q_k + \left( \frac{dQ}{dP} \Big|_{P=P_k} \right) \cdot (P_k^{MC} - P_k) \\ &\approx Q_k + \left( \frac{dQ}{dP} \Big|_{P=P_k^{MC}} \right) \cdot (P_k^{MC} - P_k). \end{aligned} \quad (22)$$

Its variation can thus be obtained

$$\sigma_{Q,k}^2 \approx \left( \frac{dQ}{dP} \Big|_{P=P_k^{MC}} \right)^2 \cdot \sigma_k^2, \quad (23)$$

where

$$\frac{dQ}{dP} \Big|_{P=P_k^{MC}} = -\frac{\sqrt{2}}{2} \pi \cdot \exp \left\{ \left[ \operatorname{erfc}^{-1} \left( 2P_k^{MC} \right) \right]^2 \right\}. \quad (24)$$

Since the BERs (and also the  $Q$  values) corresponding to different margins are estimated by independent Monte Carlo samplings, the  $K$  dimensional variable  $\mathbf{Q}^{MC} = \{Q_1^{MC}, \dots, Q_K^{MC}\}$  satisfies the following jointly Normal distribution

$$\mathbf{Q}^{MC} \sim Gauss(\boldsymbol{\mu}_Q, \boldsymbol{\Sigma}_Q), \quad (25)$$

where the mean vector  $\boldsymbol{\mu}_Q$  and variance matrix  $\boldsymbol{\Sigma}_Q$  are defined as

$$\boldsymbol{\mu}_Q = [Q_1 \quad Q_2 \quad \dots \quad Q_K]^T, \quad (26)$$

$$\boldsymbol{\Sigma}_Q = \operatorname{diag}[\sigma_{Q,1}^2 \quad \sigma_{Q,2}^2 \quad \dots \quad \sigma_{Q,K}^2]. \quad (27)$$

According to the linear assumption between  $Q$  and the margin  $m$ , the mean vector in (26) can be written as

$$\boldsymbol{\mu}_Q = \mathbf{A}\boldsymbol{\Theta}, \quad (28)$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & m_1 \\ 1 & m_2 \\ \vdots & \vdots \\ 1 & m_K \end{bmatrix} \quad (29)$$

$$\boldsymbol{\Theta} = [c \quad d]^T. \quad (30)$$

The likelihood function  $L$  is proportional to

$$L \sim \exp \left[ -(\mathbf{Q}^{MC} - \boldsymbol{\mu}_Q)^T \boldsymbol{\Sigma}_Q^{-1} (\mathbf{Q}^{MC} - \boldsymbol{\mu}_Q) \right] \quad (31)$$

The optimal value of  $\boldsymbol{\Theta} = [c \quad d]^T$  can be solved by maximizing (31).

$$\boldsymbol{\Theta} = (\mathbf{A}^T \boldsymbol{\Sigma}_Q^{-1} \mathbf{A})^{-1} \mathbf{A}^T \boldsymbol{\Sigma}_Q^{-1} \mathbf{Q}^{MC} \quad (32)$$

Eq. (32) reveals that the estimators  $Q_k^{MC}$  are weighted by the inverse of the covariance matrix (27). It means that if the variance of the estimator  $Q_k^{MC}$  is large, it becomes less important in the estimation of linear coefficients. In other words, the more sample points we collect to measure the estimator, the greater it will affect the final result. By using this MLE method, we are able to get a statistically optimal result of the linear formulation coefficients.

### B. Gamma Function Evaluation

As shown in (17), the hyper-parameter  $a$  controls our "confidence" of the prior knowledge. If  $a$  is larger, the estimated BER value is closer to the prediction by extrapolation. Otherwise, it will be close to the estimation from direct measurement. However, it's impossible to know the value of  $a$  in advance. A MLE method is used to determine the hyper-parameter  $a$  in [10], which aims to find the  $a$  value that maximize the marginal distribution:

$$\begin{aligned} p(\mathbf{x}|a) &= \int p(\mathbf{x}, \beta|a) \cdot d\beta \\ &= \frac{\Gamma((a-1)/(1-\rho_e)+2) \Gamma(N_c+a) \Gamma(N-N_c+(a-1)\rho_e/(1-\rho_e)+1)}{\Gamma(a) \Gamma((a-1)\rho_e/(1-\rho_e)+1) \Gamma(N+(a-1)/(1-\rho_e)+2)} \end{aligned} \quad (33)$$

The Gamma function in (30) is defined as

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt \quad (34)$$

The optimization problem is not essentially convex, but since it is a one-dimensional problem, it can be solved by linear search.

However, both  $N_c$  and  $N$  are very large (up to  $1e^{10}$ ) in our application. One should note that the calculation of Gamma function is quite time-consuming when  $x$  is very large.

We use Stirling's formula to deduce a much simpler approximation of the Gamma function [12]. For  $x$  with  $\operatorname{Re}(x) > 0$ , we have the following approximation of Gamma function via repeated integration by parts [13]

$$\ln(\Gamma(x)) \sim \left( x - \frac{1}{2} \right) \ln(x) - x + \frac{1}{2} \ln(2\pi) + \sum_{n=1}^{\infty} \frac{B_{2n}}{2n(2n-1)x^{2n-1}}, \quad (35)$$

where  $B_n$  is the  $n$ -th Bernoulli number. The formula is valid for  $x$  large enough in absolute value, with an error term of  $O(x^{-2m-1})$  when the first  $m$  terms are used. As a result, we can get the corresponding approximation as

$$\Gamma(x) = \sqrt{\frac{2\pi}{x}} \left( \frac{x}{e} \right)^x \left( 1 + O\left( \frac{1}{x} \right) \right) \quad (36)$$



Since  $x$  in our case is very large,  $O(1/x)$  in (36) can be eliminated and which gives the following simple approximation of gamma function

$$\Gamma(x) \sim \sqrt{2\pi} e^{-x} x^{x-\frac{1}{2}}. \quad (37)$$

By using the approximation equation (37), Gamma function in (30) can be easily and sufficiently calculated to determine the optimal value of  $a$ .

### C. Summary

Algorithm 1 summarized our proposed method for BER estimation. The final BER estimation consists of the prior knowledge by extrapolation and the observed data. In order to perform the extrapolation, we need to get data at larger margins, and then calculate the linear formula for extrapolation. After extrapolation, we use similar algorithm in [10] to estimate BER based on MAP.

#### Algorithm 1: BER estimation via Bayesian inference

1. Start from a given set of BER values at  $K$  margins  $\{P_k^{MC}, k=1 \dots K\}$
2. Calculate the value of  $Q$  by using (10).
3. Using (29) to determine the linear coefficients.
4. Performing extrapolation and inverse transformation (4) to get the prior knowledge  $\beta$ .
5. Using linear search to find the optimal value for  $a$  by (30).
6. Determine the BER value by using (14)

## V. EXPERIMENTAL RESULTS

In this section, a set of silicon measurement data from an industrial partner is used to demonstrate the efficiency of our proposed method. As described in section III and IV, our proposed algorithm is composed of two steps: BER prediction by extrapolation and BER correction by Bayesian inference. Their separate effectiveness is demonstrated in the following two subsections. All experiments are performed on a server with 2.66GHz dual-core CPU and 4GB memory.

The test data consist of 4000 groups of data. Each group provides five measurement results with different time/voltage margins. Since the real BERs are not available in these test data, in our experiments, we assume the BER of the smallest margin as the underlying “real” BERs. We use the BER values with larger margins to estimate the BER value with smallest margin. The estimated BER is then compared with the BER value with the smallest margin to judge the accuracy of estimation. For clearance, The  $Q$  values, margin values are normalized in the experiments.

### A. Bit Error Rate Prediction

The first group of measurement results shows the case that prediction is very close to the real value. As illustrated in Fig. 4(a), the transformed  $Q$  values nearly follow a linear relationship. The four points with larger margin values are used for linear fit. The linear formulation is obtained via MLE method proposed in section IV. It should be noted that the obtained line accurately approximates the relationship, and the  $Q$  value obtained by extrapolation (the point at normalized margin value 0.85) is very close to the real  $Q$  value. With

inverse transformation, the predicted BER by extrapolation is 7% larger than the “real” value.

However, as stated in section II, because the linear assumption in (3) may not be valid, extrapolation prediction may not give accurate estimation. Another group of five measurements is used to show this problem. An error between extrapolation value and real  $Q$  value can be observed in Fig. 4(b). In this case, the predicted BER by extrapolation is more than 2 times of the “real” value.

To illustrate the efficiency of our proposed MLE method, we also perform LSE to fit all the 4000 groups of data. We then compare the two sets of extrapolated value by calculating their logarithm ratio to the “real” BER value:

$$Ratio = \log_{10} \left( \frac{BER_{extrap}}{BER_{real}} \right), \quad (38)$$

The average logarithm ratio of MLE method is 0.0608, while the average ratio of LSE method is 0.0670. Since the MLE method has a logarithm mean closer to 0, it means that the extrapolation obtained by MLE method is more similar to the real value.

### B. BER Calibration by BMF

In this subsection, we use the same two groups of data in subsection A. For each case, we take the predicted BER value as the prior knowledge of BMF. We then generate random samples with their “real” BER values as the probability of taking “0”. Those random samples simulate the process of detecting bit error.

For the first example, the predicted BER by extrapolation is 7% larger than the “real” value. We implemented two methods here: (i) the traditional direct test (ii) our proposed Bayesian inference. We vary the number of random samples and calculate the corresponding estimation error defined as

$$Error = \left| \frac{BER_{esti} - BER_{real}}{BER_{real}} \right|, \quad (39)$$

where  $BER_{esti}$  is the estimated BER by direct test or our proposed method,  $BER_{real}$  is the “real” BER value.

In order to average the random fluctuations, the relative errors are calculated from 200 repeated experiments based on independent random samples. As illustrated in Fig. 5, since BER is extremely low, it is difficult for direct method to observe one bit error, and the relative error is larger than 100% when only  $10^9$  tests are available.

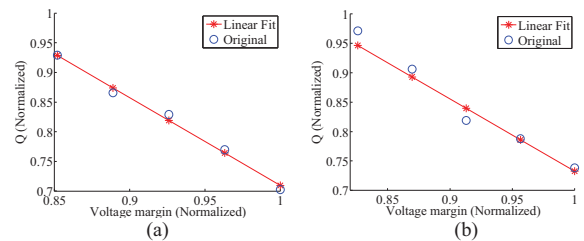


Fig. 4. The straight line obtained by MLE method and the extrapolation data

However, our proposed method can give estimation with a relative error lower than 40% with the same set of data. It is also noticed that the direct method requires more than  $8 \times 10^9$  tests to give the same accuracy. Since the time cost of BER estimation is mainly made up of bit error test, our proposed method achieves 8x runtime speed-up over direct method in this case.

The second case demonstrates the algorithm's efficiency when the prediction given by extrapolation is not accurate. The predicted BER by extrapolation here is 147% larger than the real value. We generate samples based on the "real" BER value, and then calculate relative errors by (35). The experiments are also repeated 200 times to average out the random fluctuation. Fig. 6 plots the estimation error of two methods as a function of the number of bit samples.

It can be seen from Fig. 6 that our proposed method gives much better estimation than direct method when the number of bits is only  $10^9$ . Our proposed method achieves 5x speed-up over direct method in this experiment.

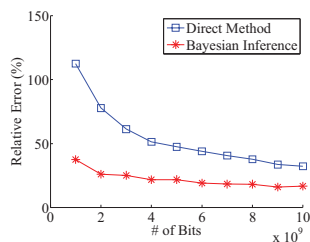


Fig. 5. The estimation error of BER is plotted as a function of the number of bits transmitted

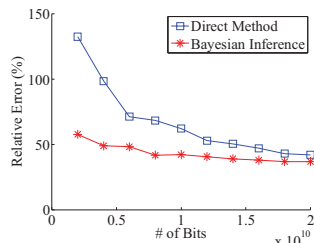


Fig. 6. The estimation error of BER is plotted as a function of the number of bits transmitted when predicted value by extrapolation is not accurate.

## VI. CONCLUSIONS

In this paper, a novel method is proposed to efficiently estimate BER by Bayesian Model Fusion. As conventional extrapolation may not be accurate and direct test is not practical, our proposed method borrows conventional extrapolated BER value as prior knowledge, and combine it with additional measurement data to estimate the BER value via Bayesian inference. In addition to applying BMF algorithm, we further develop some novel methodologies to enhance the algorithm in our BER estimation case. Experimental results demonstrate that our proposed method achieves up to 8x speed-up over direct estimation method without loss of accuracy.

## ACKNOWLEDGMENT

This research is supported partly by National Natural Science Foundation of China (NSFC) research project 61125401, 61376040, 61228401, 61474026 and 91330201, partly by the National Basic Research Program of China under the grant 2011CB309701, partly by the National Major Science and Technology Special Project of China (2011ZX01035-001-001-003, 2014ZX02301002-002), 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.

## REFERENCES

- [1] K. Cheng and H. Chang, "Recent advances in analog, mixed-signal, and RF testing," *IPSS T-SLDM*, vol. 3, pp. 19-46, Feb. 2
- [2] Stephens, Ransom. "Jitter analysis: The dual-Dirac model, RJ/DJ, and Q-scale." Agilent Technologies Whitepaper, Literature Number (2004).
- [3] National Committee for Information Technology Standardization (NCITS) (2003) T11.2/ Project 1316-DT/ Rev 10.0, Fibre Channel-Methodologies for Jitter and Signal Quality Specification-MJSQ, 10 March 2003
- [4] Hong, Dongwoo, and Kwang-Ting Cheng. Efficient test methodologies for high-speed serial links. Vol. 51. Springer, 2009.
- [5] Derickson, Dennis, and Marcus Müller. Digital communications test and measurement: high-speed physical layer characterization. Pearson Education, 2007.
- [6] X. Li, et al., "Efficient parametric yield estimation of analog/mixed-signal circuits via Bayesian model fusion," *IEEE ICCAD*, pp. 627-634, 2012.
- [7] 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.
- [8] C. Gu, et al., "Efficient moment estimation with extremely small sample size via Bayesian inference for analog/mixed-signal validation," *IEEE DAC*, 2013.
- [9] 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.
- [10] C. Fang, et al., "BMF-BD: Bayesian Model Fusion on Bernoulli Distribution for Efficient Yield Estimation of Integrated Circuits" *IEEE DAC*, 2014.
- [11] C. Bishop, "Pattern Recognition and Machine Learning," Prentice Hall, 2007.
- [12] S. Sun, et al., "Fast Statistical Analysis of Rare Circuit Failure Events via Scaled- Sigma Sampling for High-Dimensional Variation Space" *IEEE ICCAD*, 2013.
- [13] [http://en.wikipedia.org/wiki/Stirling's\\_approxim](http://en.wikipedia.org/wiki/Stirling's_approxim)