Fast and Accurate BER Estimation Methodology for I/O Links Based on Extreme Value Theory

Alessandro Cevrero[2,3], Nestor Evmorfopoulos[1], Charalampos Antoniadis[1], Paolo Ienne[2], Yusuf Leblebici[3], Andreas Burg[3] and George Stamoulis[1]
[1]Department of Computer and Communications Engineering, University of Thessaly, Volos, Greece
[2]School of Computer and Communications Science, EPF Lausanne, 1015 Lausanne, Switzerland
[3]School of Engineering, EPFL Lausanne, 1015 Lausanne, Switzerland

Abstract—**This paper introduces a novel approach towards the statistical analysis of modern high-speed I/O and similar communication links, which is capable of reliably to determine extremely low (~10^{-12} or lower) bit error rates (BER) by using techniques from *extreme value theory* (EVT). The new method requires only a small amount of voltage values at the received eye center, which can be generated by running circuit/system level simulations or measuring fabricated I/O circuits, to predict link BERs. Unlike conventional techniques, no simplifying assumptions on link noise and interference sources are required making this approach extremely portable to any communication system operating with very low BER. Our experimental results show that the BER estimates from the proposed methodology are on the same order of magnitude as traditional time domain, transient eye diagram simulations for links with BER of 10^{-6} and 10^{-5} operating at 9.6 and 10.1 Gbps respectively.**

***Index Terms*—BER, EVT, I/O Links**

I. Introduction

Driven by the use of cloud computing and data streaming, chip I/O bandwidth needs are doubling each year, reaching data rates up to 28 Gb/s in 2012 [1]. These trends are met through sophisticated architectures including complex equalization techniques in conjunction with advanced semiconductor process technology. The increasing complexity requires improvements in link modeling and development of trustable BER estimation methodology to enable optimizations and design validation. Such I/O links operate at very low BER, (< 10^{-15}) precluding the use of Monte-Carlo techniques based on several simulations of the circuit model of the link. Indeed, a probability of 10^{-15} means that roughly 10^{15} simulations are needed just to produce one erroneous bit (and many more are needed to obtain a reliable estimate of error probability). However, current simulation technology can only handle a few thousand simulations within a reasonable time, and this is not expected to improve because advances in simulation technology are offset by the increasing data-rate (as the data rate goes up, the time step required for simulation becomes smaller) and since link complexity and the need for more sophisticated circuit models also increases.

In the past, link analysis was performed using worst-case methods, such as *peak distortion analysis* (PDA) [2]. In that scenario, worst-case interference noise sources, due to *inter symbol interference* (ISI) and crosstalk, are superimposed to compute voltage margins at the sampling position. Fig. 1 depicts an eye diagram constructed by simulating 100K symbols, where also the magnitude and importance of a worst case ISI crosstalk sequence is shown as dashed line. Even though this technique is extremely useful for equalizer architectures exploration it cannot predict the BER of the system reliably.

More recently, statistical link analysis has been introduced. This technique resorts to probabilistic techniques to derive BER estimates analytically (without explicit circuit simulations) by starting with appropriate probabilistic models of all random noise and interference sources of the link, and by calculating intermediate probability distributions due to the various link components to the receiver output. However, the exact calculation of intermediate probability distributions is not possible for arbitrary and realistic probabilistic noise models, and therefore inevitably these approaches entail many assumptions and approximations both in the noise models and the calculation of intermediate distributions. A summary of previous methods is presented here.

II. Related work

Stojanovic et al. [3] proposed to compute ISI and crosstalk bounded distributions analytically. They converted TX and RX jitter noise from time domain to voltage domain (generating unbounded distributions) and convolved interference and noise sources to generate the RX sample distribution. By sweeping the sampling phase, a statistical eye contour can be constructed for a defined BER target. The shortcoming of this approach is that TX/RX jitter noise and interference sources are assumed to be statistically independent. In addition, the formulation is limited to Gaussian jitter only. Recognizing the strong interaction between these two signal impairments, Balamurugan et al. [4] proposed a segment-based analysis where data transitions are divided into segments centered on the nominal data transitions. The goal is to compute the contributions of individual segments and appropriately combine them. Moreover, they propose to use fitting methods to extract jitter information from few jitter values (roughly 200K) generated from accurate Spice-like simulations. Generally, I/O-link architects rely on the well-known dual Dirac jitter model proposed by Agilent [5], where binomial distribution is combined with two Gaussians. Although this approximation is far from describing the physical original of jitter, it is widely adopted for its tractability.

The current paper proposes an approach of statistical nature towards BER estimation, which retains the benefits of simulation that can handle any conceivable noise model, but attempts to statistically estimate very low BERs from a small and tractable sample set of only a few thousand link simulations (either system or circuit simulations). The approach is based on the field of extreme value theory (EVT) [6]-[9], which is the relevant branch of statistics for the estimation of very low probabilities which lie away from the center and into the tail of an unknown output distribution. Similar procedures based on EVT have been successfully applied in the VLSI realm for maximum power estimation [10], analog testing [11], and memory yield estimation [12].

Interestingly, the proposed BER estimation method can also be applied to I/O transceivers to perform BER analysis within seconds.

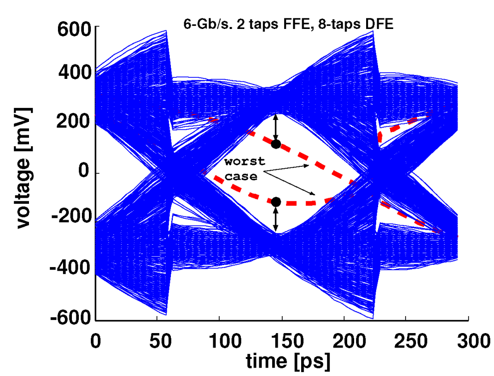


Figure 1. Eye diagram showing the importance and the magnitude of worst-case sequence computed using PDA. It could be noticed the large gap existing between the simulated eye opening and the worst case sequence.

978-3-9815370-0-0/DATE13/©2013 EDAA

III. Theoretical background

A. Definition of BER

In I/O links using a BPSK signaling format, the binary symbols '0' and '1' are mapped to voltages V_0, V_1 before being transmitted through the channel. In differential links $V_{0,1}$ is defined as the difference between the two differential lines, therefore $V_0 = +V$, while $V_1 = -V$. Due to the superposition of various random noise and interference sources along the link, the receiver (RX) voltages of symbols '0' and '1' are regarded as random variables X_- and X_+, with (unknown) *distribution functions* (df) $F_-(x)$ and $F_+(x)$, and density functions $f_-(x) = \frac{dF_-(x)}{dx}$ and $f_+(x) = \frac{dF_+(x)}{dx}$ respectively (Fig. 2). The mean of X_- represents the nominal voltage of '0' and is negative $(E[X_-] < 0)$, while the mean of X_+ represents the nominal voltage of '1' and is positive $(E[X_+] > 0)$. The bit error rate (BER) for symbol '0' is defined as the probability of X_- to be positive (in which case it is erroneously detected as '1'), i.e., $B_0 = \Pr[X_- > 0] = 1 - F_-(0)$. Likewise, the BER for '1' is defined as the probability of X_+ to be negative (in which case it is erroneously detected as '0'), i.e., $B_1 = \Pr[X_+ < 0] = F_+(0)$. For both distributions the zero voltage level lies far into the tail (right tail for '0' and left tail for '1'), where the probability is very small for the extremely low BERs needed in modern links. For this region, standard methods designed to approximate the mean are inadequate, and techniques based on EVT have to be employed instead.

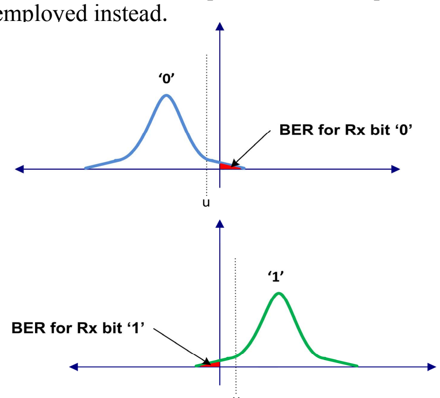


Figure 2. Probability density functions of RX voltages, and definition of bit error rate (BER) for binary symbols '0' and '1'.

B. Approximation of the tail of a probability distribution via EVT

We will describe a procedure for estimating a very low probability in the right tail of a distribution via EVT. This is the case for $B_0 = \Pr[X_- > 0]$ in our specific application, and for this purpose let us denote $X_- \equiv X$ and $F_-(x) \equiv F(x)$ for the remainder of the section. The estimation of $B_1 = \Pr[X_+ < 0]$ can be treated in a completely analogous manner by considering the distribution of $-X_+$ and estimating $\Pr[-X_+ > 0]$. As pointed out in the introduction, the objective is to estimate very low BERs on the basis of a relatively small sample set of simulated RX voltages. Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a random sample set of RX voltages corresponding to '0' (for the estimation of B_1 from a sample set $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ of '1' voltages, just consider the sample $-\mathbf{X} = (-X_1, -X_2, \ldots, -X_n)$). By "random" here we mean that the units $X_1, X_2, \ldots, X_n$ constitute *independent and identically distributed* (iid) random variables with df $F_-(x) \equiv F(x)$ each (the way to form such a random sample by *n* simulations of the link will be discussed in Section III). The units $X_1, X_2, \ldots, X_n$ of $\mathbf{X}$ can be sorted in ascending order as $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$, and the *i*-th unit $X_{i:n}$ of this sequence is called the *i*-th *order statistic* of $\mathbf{X}$. Let us suppose that the *k* upper order statistics $X_{n-k+1:n}, X_{n-k+2:n}, \ldots, X_{n:n}$ belong to the right tail of the distribution (we discuss the appropriate selection of their number *k* in Section II-D). In other words, let the $(n-k)$-th order statistic $X_{n-k:n}$ constitute a high threshold *u* marking the beginning of the right tail. The sample set $\mathbf{X}_{ex} = (X_{n-k+1:n}, X_{n-k+2:n}, \ldots, X_{n:n})$ of the order statistics which are larger than $u \equiv X_{n-k:n}$ is called the sample of *exceedances* over threshold *u*. It is not difficult to infer that this sample set follows a distribution $F_u(x)$:

$$(1)\quad F_u(x) = \Pr[X \le x \mid X > u] = \frac{\Pr[X \le x, X > u]}{\Pr[X > u]} = \frac{F(x) - F(u)}{1 - F(u)}, \quad x \ge u$$

Now, a fundamental result from EVT [7] states that *irrespective* of $F(x)$ and under conditions normally satisfied in practice, the exceedance df $F_u(x)$ approaches asymptotically (i.e., for large enough *u*) a *generalized Pareto* (GP) distribution $G_{\beta,\gamma}(x)$ with the following form:

$$(2)\quad G_{\beta,\gamma}(x) = 1 - \left(1 - \gamma \frac{x-u}{\beta}\right)^{1/\gamma}, \quad \begin{array}{ll} u \le x \le +\infty & \text{if } \gamma \le 0 \\ u \le x < u + \beta/\gamma & \text{if } \gamma > 0 \end{array},$$

where $\beta > 0$ and $\gamma \in \Re$ are parameters and $u \equiv X_{n-k:n}$ is the chosen constant threshold. The parameter β is a *scale* parameter that characterizes the spread of the distribution around its mean, and roughly corresponds to the standard deviation parameter σ of the normal distribution. The parameter γ is called *shape* parameter or *tail index*, and is more important since it is connected with the right tail of the parent distribution $F(x)$. If $\gamma < 0$ then $F(x)$ has a long (or *heavy*) right tail and the associated random variable X is unbounded from above (i.e., $\sup\{x : F(x) < 1\} = +\infty$). If $\gamma > 0$ then $F(x)$ has a short right tail and the random variable X is upper bounded (i.e., $\sup\{x : F(x) < 1\}$ is finite). The case $\gamma = 0$ is interpreted as $\gamma \to 0$ [reducing the GP to the *exponential* distribution $G_{\beta,0}(x) = 1 - \exp(-x/\beta)$] and indicates that $F(x)$ has a moderate right tail.

If we can fit the GP distribution (2) to the sample of exceedances $\mathbf{X}_{ex}$ (i.e., select appropriate values $\hat{\beta}$ and $\hat{\gamma}$ of the parameters β and γ) so as to approximate $F_u(x)$ by $G_{\hat{\beta},\hat{\gamma}}(x)$, and also approximate $1 - F(u) = 1 - F(X_{n-k:n}) = \Pr[X > X_{n-k:n}]$ by the percentage k/n of upper order statistics in total sample size, we can estimate the unknown $F(x)$ for $x \ge u$ (i.e., for points further into the tail) from (1) and (2) as:

$$(3)\quad F(x) = (1 - F(u))F_u(x) + F(u) = 1 - (1 - F(u))(1 - F_u(x))$$
$$\approx 1 - \frac{k}{n}\left(1 - \hat{\gamma}\frac{x - X_{n-k:n}}{\hat{\beta}}\right)^{1/\hat{\gamma}}$$

and finally the desired $B_0 = 1 - F(0)$ as:

$$(4)\quad \hat{B}_0 \approx \frac{k}{n}\left(1+\hat{\gamma}\frac{X_{n-k:n}}{\hat{\beta}}\right)^{1/\hat{\gamma}}$$

The next two sub-sections discuss the pending issues of finding appropriate estimates $\hat{\beta},\ \hat{\gamma}$ of the GP parameters $\beta,\ \gamma$ and setting an appropriate number k of order statistics belonging to the right tail.

C. *Estimation of parameters of generalized pareto distribution*

Maximum likelihood (ML) estimation is the standard way of estimating the unknown parameters of a distribution on the basis of a sample set. It amounts to maximizing the joint density function of the iid units of the sample w.r.t. the unknown parameters (i.e., the *likelihood* function), or more typically its natural logarithm (known as the *log-likelihood* function). The density function of the GP distribution (2) is

$$(5)\quad g_{\beta,\gamma}(x)=\frac{dG_{\beta,\gamma}(x)}{dx}=\frac{1}{\beta}\left(1-\gamma\frac{x-u}{\beta}\right)^{(1/\gamma)-1},\quad \begin{array}{ll} u\le x\le+\infty & \text{if}\quad \gamma\le 0\\ u\le x<u+\beta/\gamma & \text{if}\quad \gamma>0\end{array}$$

and the corresponding log-likelihood function for the sample $\mathbf{X}_{ex}$ is

$$(6)\quad \log L(\beta,\gamma)=\log\prod_{i=1}^{k}\left(\frac{1}{\beta}\left(1-\gamma\frac{X_{n-i+1:n}-X_{n-k:n}}{\beta}\right)^{(1/\gamma)-1}\right)$$
$$=-k\log\beta+\left(\frac{1}{\gamma}-1\right)\sum_{i=1}^{k}\log\left(1-\gamma\frac{X_{n-i+1:n}-X_{n-k:n}}{\beta}\right).$$

The maximization of the above bi-variate function w.r.t. (β,γ) to obtain the ML estimates $(\hat{\beta},\hat{\gamma})$ can be performed by numerical routines for multivariable nonlinear optimization. More effectively, we can perform a reparameterization $(\beta,\gamma)\to(\tau,\gamma)$ with $\tau\equiv\gamma/\beta$, and differentiate Eq. (6) w.r.t. τ and γ. The estimate $\hat{\tau}$ can then be found via numerical solution of the following one-dimensional equation in the interval $\tau\in(-\infty,(X_{n:n}-u)^{-1})$ (see [13] for details)

$$(7)\quad \frac{1}{\tau}+\left(\frac{1}{\sum_{i=1}^{k}\log\left(1-\tau(X_{n-i+1:n}-X_{n-k:n})\right)}+\frac{1}{k}\right)\sum_{i=1}^{k}\frac{X_{n-i+1:n}-X_{n-k:n}}{1-\tau(X_{n-i+1:n}-X_{n-k:n})}=0$$

and the estimate $\hat{\gamma}$ (and $\hat{\beta}=\hat{\gamma}/\hat{\tau}$) can be computed by

$$(8)\quad \hat{\gamma}=-\frac{1}{k}\sum_{i=1}^{k}\log\left(1-\hat{\tau}(X_{n-i+1:n}-X_{n-k:n})\right).$$

The ML estimation method has also the benefit of providing confidence intervals for the estimates $(\hat{\beta},\hat{\gamma})$ as well as the derived estimate $\hat{B}_0$. Specifically, provided that $\gamma<1/2$ (the converse case $\gamma\ge 1/2$ is very rare and corresponds to distributions with very short tails and a right endpoint very close to their mean), the ML estimates $(\hat{\beta},\hat{\gamma})$ are known to be jointly normally-distributed with means (β,γ) and the following covariance matrix [14] (where X is any GP-distributed random variable with density function $g_{\beta,\gamma}(x)$):

$$\mathbf{V}=-\frac{1}{k}\left[\begin{array}{cc} E\left\{\frac{\partial^2\log g_{\beta,\gamma}(X)}{\partial\beta^2}\right\} & E\left\{\frac{\partial^2\log g_{\beta,\gamma}(X)}{\partial\beta\partial\gamma}\right\}\\ E\left\{\frac{\partial^2\log g_{\beta,\gamma}(X)}{\partial\beta\partial\gamma}\right\} & E\left\{\frac{\partial^2\log g_{\beta,\gamma}(X)}{\partial\gamma^2}\right\}\end{array}\right]^{-1}\Bigg|_{(\beta,\gamma)=(\hat{\beta},\hat{\gamma})}$$
$$=\frac{1}{k}(1-\hat{\gamma})\begin{bmatrix}2\hat{\beta}^2 & \hat{\beta}\\ \hat{\beta} & 1-\hat{\gamma}\end{bmatrix}.$$

The estimate $\hat{\kappa}\equiv k/n$ of $1-F(X_{n-k:n})$ in (4) has also a variance that equals $\mathrm{var}(\hat{\kappa})=k/n^2$ (provided that $k<<n$) [8], and is *independent* of the ML estimates $(\hat{\beta},\hat{\gamma})$ (since it is known that the number of exceedances k is independent of their numerical values upon which $(\hat{\beta},\hat{\gamma})$ are estimated [15]). If we denote

$$\mathbf{d}=\begin{bmatrix}\frac{\partial\hat{B}_0}{\partial\hat{\beta}}\\ \frac{\partial\hat{B}_0}{\partial\hat{\gamma}}\end{bmatrix}=-\underbrace{\frac{k}{n}\left(1+\hat{\gamma}\frac{X_{n-k:n}}{\hat{\beta}}\right)^{1/\hat{\gamma}}}_{\hat{B}_0}\begin{bmatrix}\frac{1}{\hat{\beta}}\frac{X_{n-k:n}}{\hat{\gamma}X_{n-k:n}+\hat{\beta}}\\ \frac{1}{\hat{\gamma}}\left(\frac{1}{\hat{\gamma}}\log\left(1+\hat{\gamma}\frac{X_{n-k:n}}{\hat{\beta}}\right)-\frac{X_{n-k:n}}{\hat{\gamma}X_{n-k:n}+\hat{\beta}}\right)\end{bmatrix}.$$

the variance of the BER estimate, (4) can be found by applying the so-called "delta" method [16] as

$$(9)\quad \mathrm{var}(\hat{B}_0)=\mathbf{d}^T\mathbf{V}\mathbf{d}+\left(\frac{\partial\hat{B}_0}{\partial\hat{\kappa}}\right)^2\mathrm{var}(\hat{\kappa})$$
$$=\frac{1}{k}\hat{B}_0^2\left[\left(\left(2-\frac{1}{\hat{\gamma}}\right)\frac{X_{n-k:n}}{\hat{\gamma}X_{n-k:n}+\hat{\beta}}+\left(\frac{1}{\hat{\gamma}}-1\right)\frac{1}{\hat{\gamma}}\log\left(1+\hat{\gamma}\frac{X_{n-k:n}}{\hat{\beta}}\right)\right)^2+(1-2\hat{\gamma})\left(\frac{X_{n-k:n}}{\hat{\gamma}X_{n-k:n}+\hat{\beta}}\right)^2+1\right].$$

and the $(1-\delta)\times 100\%$ confidence interval (corresponding to confidence level $1-\delta$) ultimately results in

$$(10)\qquad \left|B_0-\hat{B}_0\right|\le z_{\delta/2}\sqrt{\mathrm{var}(\hat{B}_0)}\ ,$$

where $z_{\delta/2}$ is the $(\delta/2)$ -quantile of the standard normal distribution $N(0,1)$.

However, it must be remarked that the normal approximation assumed by the delta method for the sampling distribution of the estimate of Eq. (4) can be rather poor, and as such the above confidence interval is not always particularly small. A better confidence interval can often be constructed on the basis of the *profile likelihood ratio* test statistics [17], or even sometimes by adopting other parameter estimation methods such as the method of *probability-weighted moments* [18] or the *elemental percentile method* [19].

D. *Selection of tail size*

An important question in tail estimation via EVT concerns the appropriate number k of upper order statistics that are assumed to belong to the tail. In determining k we are faced with a tradeoff between variance and bias. ML estimation of $\beta,\ \gamma$ on the sample $\mathbf{X}_{ex}$ requires k to be large so that the estimates $\hat{\beta},\ \hat{\gamma}$ (and the corresponding BER estimate) are accurate, i.e., they have small variance. A too large k, however, may incorporate units that do not belong to the tail, and therefore cause $\mathbf{X}_{ex}$ to deviate from its asymptotic GP distribution, i.e., introduce bias. In practice, a rule of thumb that is often used is to set k at the upper 10% of the sample set $\mathbf{X}$, i.e., fix $k=0.1n$. However, this choice is totally arbitrary and a much better strategy is to perform ML estimation of the tail index γ

(which ought to be independent of the selection of k) for a whole range of values between some $k_{\min}$ (or $k=1$) and $k_{\max}$, and seek the optimum k^* inside a region, where the estimated $\hat{\gamma}$ is stable.

A graphical procedure can be adopted for this by drawing a plot of $\hat{\gamma}$ versus k. The plot is expected to exhibit a "*plateau*" that is surrounded by strong fluctuations of $\hat{\gamma}$ on the left (due to large variance for smaller k) and a gradual shift from the stable $\hat{\gamma}$ on the right (due to bias for larger k). The optimum k^* is the largest value of k within this plateau. A heuristic to find the right endpoint of a plateau, which works reasonably well in practice (but is still advisable to be accompanied by visual judgment), consists of choosing k^* as the value between $k_{\min} \le k \le k_{\max}$ where the following function is minimized [6]

$$(11) \qquad p(k) = \frac{1}{k}\sum_{i=1}^{k}\left|\hat{\gamma}_i - \mathrm{med}(\hat{\gamma}_1,\ldots,\hat{\gamma}_k)\right|,$$

where the estimate $\hat{\gamma}_i$ of γ is based on the i upper order statistics, and $\mathrm{med}(\cdot)$ denotes the median value.

More rigorous but more complicated procedures for automatic selection of k can be found in [20] – [21]. We also note that there is another graphical procedure for tail size selection called the *mean excess plot* [9], which, however, has a greater degree of subjectivity and does not lend itself to a good heuristic for automatic size selection.

IV. Proposed Methodology

The first stage in the proposed statistical approach for BER estimation is the creation of the random sample set of RX voltages from n simulations of the link. This entails a circuit-level simulator or an appropriate circuit model of the link. It also requires appropriate realistic probabilistic models for all random noise contributions in the link. If there are s random noise sources $Y_1, Y_2, \ldots, Y_s$ (not necessarily independent), the specification of the probabilistic model of their combined random vector $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_s)$ effectively amounts to characterizing its joint (multivariate) distribution function $H(y_1, y_2, \ldots, y_s) \equiv H(\mathbf{y})$. As already mentioned in the introduction, the proposed approach does not impose any restrictions, neither on the link circuit model nor on the probabilistic noise models. The only real requirement is to ensure that the sample set of RX voltages, upon which statistical BER estimation is to be carried out, comprises of independent and identically distributed (iid) units from the unknown target distribution. This is simply achieved by *randomly* generating n independent noise vectors $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n$ from their joint distribution $H(\mathbf{y})$, and entering each specific vector into the circuit simulator with the link model (e.g., a SPICE netlist), in order to perform n simulations for the RX output. The result will be a random sample set $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ with iid units from the unknown distribution of the RX output voltage.

Regarding the sample size n, we have chosen a number of $n = 20000$ units to conduct our experiments (presented in SectionV). Assuming that the tail comprises a portion of about 10% of the sample, this provides a tail size of approximately $k \approx 2000$ units (upper order statistics) to fit the GP parametric model, which is generally more than adequate for statistical estimation. Of course, if we have the simulation capacity, we can increase the sample size (along with the tail size) in order to improve the estimation accuracy. As also pointed out in Section II-D, it is not recommended to rely on a fixed tail size (like $k = 0.1n$ units) for tail estimation, but rather conduct estimation for a whole range of k values and choose the best among them. A search within a range of $k_{\min} = 0.005n$ to $k_{\max} = 0.15n$ (i.e. $k_{\min} = 100$ to $k_{\max} = 3000$ for $n = 20000$) has proven to be sufficient in most practical cases.

After the n link simulations have been executed and the random sample set $\mathbf{X}$ of RX voltages has been assembled, the second stage of the proposed approach is the actual BER estimation via EVT on the basis of the background developed in Section II. The resulting procedure is as follows:

Step 0. If the sample **X** corresponds to symbol '1', revert it as $-X=(-X_1,-X_2,\ldots,-X_n)$ (leave **X** as is, if it corresponds to symbol '0').

Step 1. Sort the units $X_1,X_2,\ldots,X_n$ of **X** in ascending order as $X_{1:n} \le X_{2:n} \le \ldots \le X_{n:n}$.

Step 2. For k=1 to $k_{\min}$-1:

- **Step 2.1.** Group together the ordered units $X_{n-k+1:n}, X_{n-k+2:n}, \ldots, X_{n:n}$ that are larger than $X_{n-k:n}$ as sample X_{ex} of exceedances over threshold $u \equiv X_{n-k:n}$.
- **Step 2.2.** Solve (7) w.r.t. τ and obtain estimate $\hat{\tau}_k$ (for the current k).
- **Step 2.3.** Calculate estimate $\hat{\gamma}_k$ of the tail index γ (for the current k) from (8).

Step 3. For $k=k_{\min}$ to $k_{\max}$:

- **Step 3.1.** Group together the ordered units $X_{n-k+1:n}, X_{n-k+2:n}, \ldots, X_{n:n}$ that are larger than $X_{n-k:n}$ as sample X_{ex} of exceedances over threshold $u \equiv X_{n-k:n}$.
- **Step 3.2.** Solve (7) w.r.t. τ and obtain estimate $\hat{\tau}_k$ (for the current k).
- **Step 3.3.** Calculate estimate $\hat{\gamma}_k$ of the tail index γ (for the current k) from (8).
- **Step 3.4.** Evaluate p(k) of (12) and check whether it is smaller than the current. If so, set $k^* \equiv k$.

Step 4. For the resultant optimum k^*, retrieve $\hat{\gamma}_{k^*} \equiv \hat{\gamma}$ and $\hat{\tau}_{k^*} \equiv \hat{\tau}$ and calculate $\hat{\beta} = \hat{\gamma}/\hat{\tau}$.

Step 5. Calculate the estimate $\hat{B}_0$ of B_0 from (4) and its confidence interval from (11).

V. Experimental Results

A. Behavioural simulations

For the purposes of this evaluation, we have developed a system-level time-domain simulator to generate RX samples which are later used for the fitting procedure. It should be noticed that in absence of measured data from fabricated prototypes, assumptions have been made for noise source distributions to generate time domain data. However, this has no impact on the final result. The simulation framework was written using MATLAB, due to the small run time required to generate a large number of samples, and due to the simplicity and ease of creating different system configurations, by sweeping filter taps and different data rates (it should be noted that in principle, a circuit simulator would be able to generate the same data allowing to model more accurately noise sources in the system). The modeled I/O link is depicted in Fig. 3. The TX features a *feed forward equalizer* (FFE) for precursor equalization, while post cursor components are removed on RX side using a *decision feedback equalizer* (DFE). *A pseudorandom bit generator* (PRBS) creates random data vectors for time domain simulations. The channel, whose frequency response is depicted in Fig. 4(a) is an 8-lanes parallel bus, therefore, *far end crosstalk* (FEXT) components are included in the simulation. Fig. 4(b) depicts the forward and FEXT pulse responses for a 166ps wide pulse. Cursor, precursor and postcursor components are clearly visible. Since this pulse response has been sampled at the RX input, precursors have been effectively removed by TX-FFE. Apart from interference sources, unbounded noises sources have been included in the simulation. TX and RX jitter samples, considered to be white and independent are modeled with dual Dirac distribution (5% UI deterministic jitter and 1% UI RMS value for random component), capturing both deterministic and random jitter components. 5 MHz sinusoidal jitter is superimposed. TX/RX jitter samples are transformed into equivalent noise at the slicer input using the methodology proposed in Stojanovic et al. [3]. Since the simulated link is assumed to be source synchronous, the effective jitter seen at RX is the difference between TX and RX jitter.

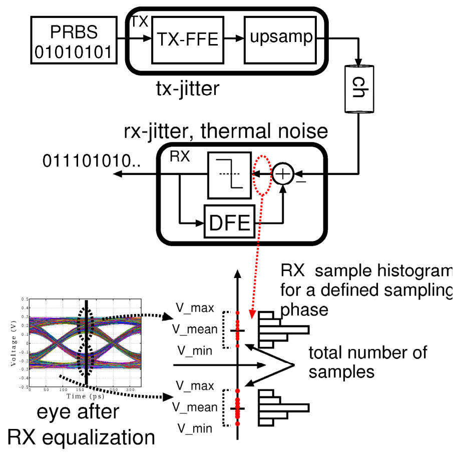
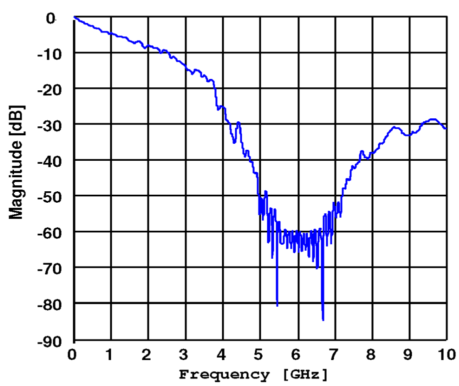
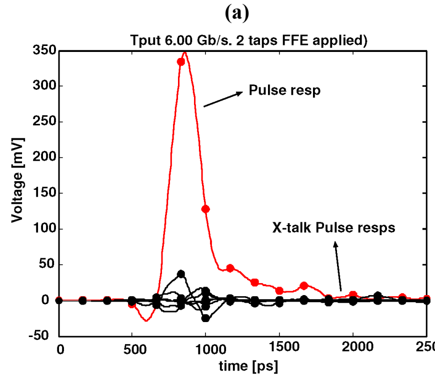


Figure 3. Block diagram of the I/O link used for time domain simulation. The RX samples used for the fitting procedure are depicted. The probabilistic RX eye can be generated by sweeping the sampling phase.



(a)

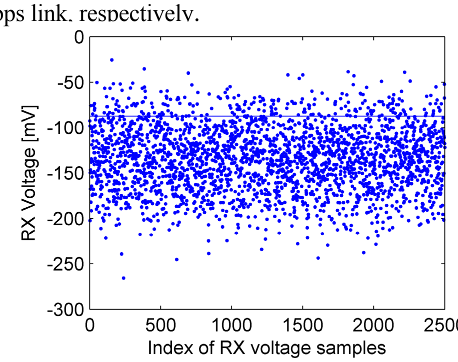


(b)

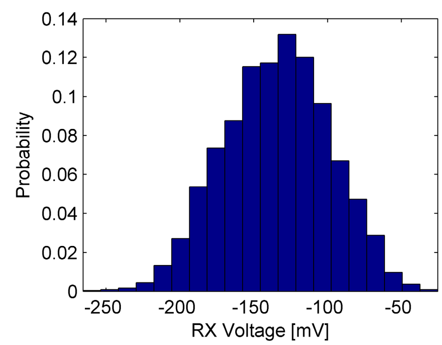
Figure 4. **(a)** Frequency Response (S21) of server backplane channel. The notches in the response are due to reflections. **(b)** Forward and croostalk pulse response for a 166 ps pulse, corresponding to 6Gb/s symbol rate.

B. BER estimation

We evaluated our proposed methodology on three samples of 20,000 trials for each symbol '0', '1' on two different data links with bandwidths of 9.6 and 10.1 Gbps. The tail of the empirical distribution function of these samples can be approximated by a GP distribution as shown in Figures 5 and 6. In order to find the k^* we evaluated the ML estimate of index γ for different tail sizes k as shown in Fig. 7(a) and then used the metric p(k). Finally we evaluated the BER for the k^* as shown in Fig. 7(b). Our experimental results in Table I show that the BER estimates are on the same order of magnitude as Monte Carlo estimates from brute-force simulations: 10^{-6} and 10^{-5} for 9.6 Gbps and 10.1 Gbps link, respectively.

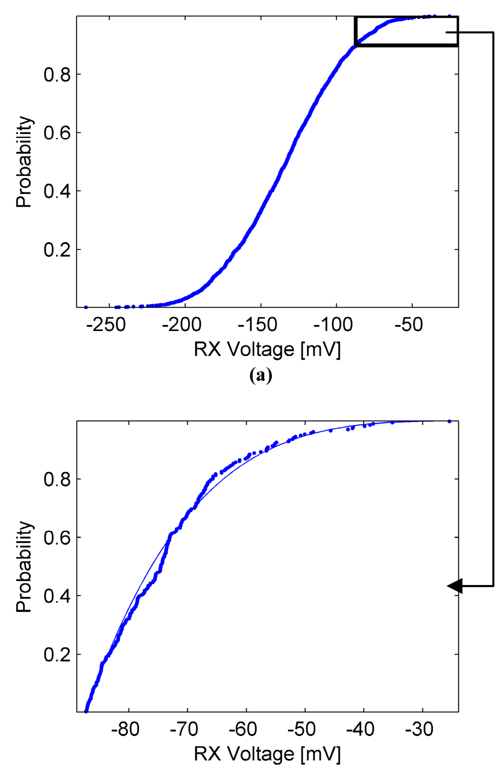


(a)



(b)

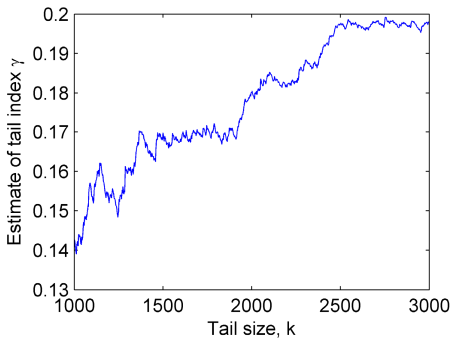
Figure 5. (**a**) Data plot and (**b**) histogram for a sample of n=2500 units corresponding to symbol '0' in the 9.6 Gbps link. The data plot also shows the exceedances over the threshold $u \equiv X_{0.9n:n}$.



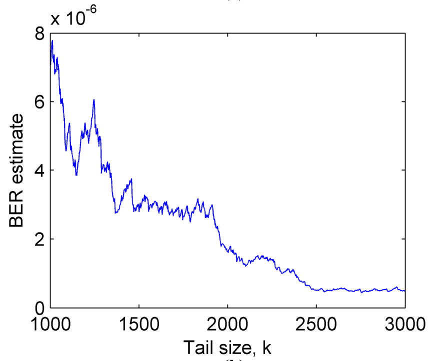
(a)

(b)

Figure 6. **(a)** Empirical distribution function (df) for the sample of Fig. 5, and **(b)** exceedance df over the threshold $u=X_{0.9n:n}$ (together with the fitted GP distribution, after ML estimation of parameters β, γ). The reader can notice how great is the fitting of a GP distribution in the tail of the empirical distribution function.



(a)



(b)

Figure 7. Plots of the **(a)** estimates of the tail index γ and **(b)** the BER for symbol '0' as function of the tail size k, for a sample of n=20,000 units in the 9.6 Gbps link. The tail size chosen by the metric p(k) was k^*=1954 upper order statistics (at the right endpoint of the perceived plateau), and the BER estimate for k* was $\hat{B}_0 = 2.10\times10^{-6}$.

TABLE I. BER ESTIMATES FOR SYMBOLS '0' AND '1', ALONG WITH 90% CONFIDENCE INTERVALS AND SELECTION OF OPTIMUM TAIL SIZE, FOR 3 DIFFERENT SAMPLES OF 20,000 UNITS FROM THE 9.6 GBPS AND 10.1 GBPS LINKS.

Link data rate	Binary symbol	Sample index	Tail size selection	BER estimate	90% conf. interval
9.6 Gbps	'0'	#1	1244	2.95×10^{-6}	7.56×10^{-6}
		#2	1954	2.10×10^{-6}	5.64×10^{-6}
		#3	1434	5.96×10^{-6}	11.9×10^{-6}
	'1'	#1	1496	4.13×10^{-6}	8.98×10^{-6}
		#2	1582	3.70×10^{-6}	8.33×10^{-6}
		#3	1790	2.64×10^{-6}	6.66×10^{-6}
10.1 Gbps	'0'	#1	2668	4.47×10^{-5}	4.35×10^{-5}
		#2	2304	4.40×10^{-5}	4.47×10^{-5}
		#3	2112	4.09×10^{-5}	4.26×10^{-5}
	'1'	#1	2140	4.90×10^{-5}	4.76×10^{-5}
		#2	2724	4.25×10^{-5}	4.25×10^{-5}
		#3	2026	4.71×10^{-5}	4.75×10^{-5}

VI. CONCLUSION

In this paper we combine the advantage of probabilistic modeling and the accuracy of detailed time domain simulations to generate reliable estimates of very low BERs for a I/O link architectures. The proposed methodology directly estimates bit error probability from a small set of accurate voltage samples observed at the receiver, rather than attempting to model all deterministic and random noise sources in the system in an analytically trackable and therefore often overly simplistic way. As a result, given the absence of assumptions concerning the origins of noise, the method is applicable to any link architecture and to complex scenarios.

REFERENCES

[1] Bulzacchelli, J et al. "A 28Gb/s 4-tap FFE/15-tap DFE serial link transceiver in 32nm SOI CMOS technology," IEEE Solid-State Circuits Conference (ISSCC), 2012.

[2] Casper, B.K.; Haycock, M.; Mooney, R.; , "An accurate and efficient analysis method for multi-Gb/s chip-to-chip signaling schemes," Symposium on VLSI Circuits, 2002.

[3] Stojanovic, V.; Horowitz, M.; , "Modeling and analysis of high-speed links," IEEE Custom Integrated Circuits Conference (CICC), 2003.

[4] Balamurugan, G.; Casper, B.; Jaussi, J.E.; Mansuri, M.; O'Mahony, F.; Kennedy, J, "Modeling and Analysis of High-Speed I/O Links," Advanced Packaging, IEEE Transactions on , vol.32, no.2, pp.237-247, May 2009.

[5] R. Stephens, "Dual Dirac Approx: Jitter Analysis: The Dual-Dirac Model, RJ/DJ, and Q-scale", 2004.

[6] R.-D. Reiss and M. Thomas, Statistical Analysis of Extreme Values, 3rd ed., Birkhauser, 2007.

[7] Beirlant, Y. Goegebeur, J. Teugels, and J. Segers, Statistics of Extremes: Theory and Applications, Wiley, 2004.

[8] S. Coles, An Introduction to Statistical Modeling of Extreme Values, Springer, 2001.

[9] P. Embrechts, C. Kluppelberg, and T. Mikosh, Modelling Extremal Events for Insurance and Finance, Springer, 1997.

[10] N. Evmorfopoulos, G. Stamoulis, and J. Avaritsiotis, "A Monte Carlo approach for maximum power estimation based on extreme value theory", IEEE Trans. Computer-Aided Design, vol. 21, no. 4, pp. 415-432, 2002.

[11] H. Stratigopoulos, "Test metrics model for analog test development", IEEE Trans. Computer-Aided Design, vol. 31, no. 7, pp. 1116-1128, 2012.

[12] A. Singhee and R. Rutenbar, "Statistical blockade: very fast statistical simulation and modeling of rare circuit events and its application to memory design", IEEE Trans. Computer-Aided Design, vol. 28, no. 8, pp. 1176-1189, 2011.

[13] A. Davison and R. Smith, "Models for exceedances over high thresholds", J. Royal Statist. Society B, vol. 52, no. 3, pp. 393-442, 1990.

[14] R. Smith, Estimating tails of probability distributions, Ann. Statist., vol. 15, no. 3, pp. 1174-1207, 1987.

[15] M. Leadbetter, "On a basis for 'peaks over threshold' modeling", Statist. Probab. Letters, vol. 12, pp. 357-362, 1991.

[16] C. Rao, Linear Statistical Inference and its Applications, 2nd ed., Wiley, 1973.

[17] O. Barndorff-Nielsen and D. Cox, Inference and Asymptotics, Chapman and Hall, 1994.

[18] J. Hosking and J. Wallis, "Parameter and quantile estimation for the generalized Pareto distribution", Technometrics, vol. 29, pp. 339-349, 1987.

[19] E. Castillo and A. Hadi, "Fitting the generalized Pareto distribution to data", J. Amer. Statist. Assoc., vol. 92, pp. 1609-1620, 1997.

[20] H. Drees and E. Kaufmann, "Selecting the optimal sample fraction in univariate extreme value estimation", Stochastic Processes and their Applications, vol. 75, pp. 149-172, 1998.

[21] A. Guillou and P. Hall, "A diagnostic for selecting the threshold in extreme value analysis", J. Royal Statist. Society B, vol. 63, pp. 293-305, 2001.