# Variational Inference approach to Jitter decomposition in High-speed Link

Bobi Shi, Thong Nguyen and Jose Schutt-Aine
Department of Electrical and Computer Engineering.
University of Illinois Urbana - Champaign
{bobishi2, tnnguye3, jesa} @illinois.edu

***Abstract*—Jitter, a timing deviation from the ideal edge position, is an unwanted phenomenon in high-speed link systems. Decomposing jitter into its components and identifying each type of jitter are beneficial to diagnose the root causes of jitter, thereof improving the system design. In recent years, variational bayes inference (VBI) has made substantial progress towards improving the efficiency of statistical modeling. This paper proposes a jitter approximation method using VBI. Applications approximating different mixtures of well-known components of jitter show good results. The approximated distribution is much closer to the true jitter distributions as compared to traditional methods.**

***Index Terms*—jitter decomposition, high-speed link system, Gaussian mixture model, stochastic variontional inference**

## I. INTRODUCTION

Over the past decades, signal analysis has been aggressively pursued in order to meet urgent demands for multi-gigabit data rate transfer. As data rates keep increasing, jitter, defined as the deviation of timing edges from their ideal positions, becomes a crucial element affecting the performance of overall high-speed link systems. In the high-speed link system, a small timing deviation might be a significant portion of the signal interval because of the fast transition and short unit interval. Then the jitter-induced errors corrupt the signals and even the clock, causing failure of the proper signal transmission. As a consequence, tighter control over jitter for high-speed link systems is needed to prevent signal failure or non-ideality, and understanding amount of jitter generated by different jitter sources is vital for high-speed link system designer to meet the requirements.

Generally, an observed total jitter can be classified into two categories: deterministic jitter (DJ) and random jitter (RJ). Deterministic jitter follows the bounded distribution and is normally separated into various components to investigate different root causes. It is further divided into periodic jitter (PJ), data-dependent jitter (DDJ) and bounded uncorrelated jitter (BUJ). PJ repeats in a sinusoidal fashion and external deterministic noise sources such as switching power supply noise can lead to PJ. DDJ is the data pattern related to jitter. The rest of the bounded jitter goes into the BUJ category. Random jitter is unpredictable jitter and is caused by unbounded jitter sources. Random jitter normally has a Gaussian distribution model because the sources of random jitter are thermal noise, 1/f flicker noise or shot noise, which fit to a bell curve.

In order to separate jitter into its components, various methods have been developed. The tail-fitting algorithm is one of the prominent jitter separation method [1]. Two Gaussian distribution curves are found to fit the tails of total jitter probability distribution. Then, the quantity of RJ and DJ can be determined through tail-fitting. However, the drawback is that a large amount of jitter samples is required to identify the fit tail part of the distribution. In [2] and [3], a better jitter decomposition method based on Gaussian mixture model (GMM) was promoted. For comparison, in this preliminary work on jitter extraction using stochastic variation inference (SVI) framework, we aslo implemented a GMM model but reformulated the maxmimum likelihood problem into a SVI problem. Convergence rate is about the same between the two methods but accuracy is improved by using SVI. In addition, rather than restricting to a limited class of models that allows an analytical solution, SVI provides a general tool to define any statistical model,regardless of the complexity.

In the next section, the Dual-Dirac model of jitter is reviewed at first. Section III presents the GMM and variational GMM algorithm to segregate the total jitter. Examples of decomposing a mixture of RJ and DJ jitters with different type of DJ are presented in Section IV. Section V concludes the paper.

## II. DUAL-DIRAC MODEL OF JITTER

As mentioned in Section I, jitter can be mainly classified into RJ and DJ, Dual-Dirac model [4] assumes that the total jitter can be explained by 2 components whose probability density functions (pdf) are

$$\begin{aligned} p\left(DJ\right) &= 0.5\delta\left(t-\mu_L\right)+0.5\delta\left(t-\mu_R\right) \\ p\left(RJ\right) &= \mathcal{N}\left(0,\sigma^2\right) \end{aligned} \tag{1}$$

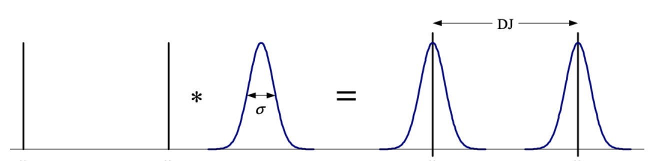


Fig. 1. Dual-Dirac model.

Illustration of total jitter distribution is shown in Figure 1. Fitting jitter data to dual-dirac model means finding $\sigma$, $\mu_L$ and $\mu_R$ to maximize the likelihood observing the data. In most cases, PJ, an example of whose distribution is shown

978-1-7281-6161-7/20/$31.00 

in Figure 2, can also be approximated by 2 Dirac functions as well.

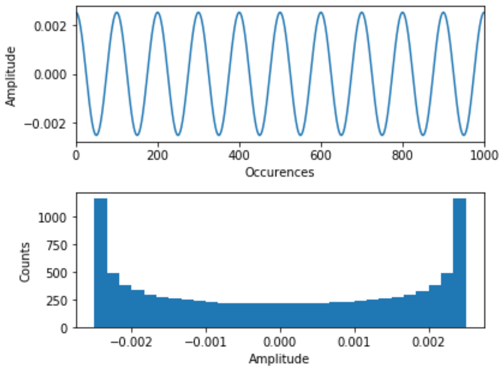


Fig. 2. Periodic jitter.

An advantage of the dual-dirac model is that it is simple and analytical. However, the trade-off is that the accuracy is low and jitter decomposed by the dual-dirac model can only be separated into either DJ or RJ, which is not very useful for debugging the circuit of interest.

## III. Statistical modeling of Jitter with GMM

Due to the oversimplicity of dual-dirac model, many attempts to improve jitter approximation have been proposed in the literature. Among those, standing out is the work in [2], a GMM is used to approximate the total jitter distribution. It is useful because GMM is a powerful and flexible model that can accurately approximate a smooth pdf.

Consider a statistical model with the latent variable $\boldsymbol{z}$, observed variable $x$, let $\boldsymbol{\theta}$ denote the set of parameters. The log-likelihood of observing a set of data $X$ is

$$\log p\left(X|\,\boldsymbol{\theta}\right) \;=\; \int \log p\left(X|\,\boldsymbol{z},\boldsymbol{\theta}\right) p\left(\boldsymbol{z}|\,\boldsymbol{\theta}\right) \mathrm{d}\boldsymbol{z} \qquad (2)$$

Statistical inference involves finding the maximum likelihood (ML) solution, $\boldsymbol{\theta}_{ML}$, that maximizes the (log-)likelihood in (2). The expectation maximization (EM) algorithm is an iterative algorithm converging the $\boldsymbol{\theta}$ to the ML solution and is especially useful when no closed-form solution is available.

A GMM of $n$ clusters is formulated using a latent variable $\boldsymbol{z}$ representing the Gaussian cluster from which a sample $\boldsymbol{x}$ comes from. Thus, $\boldsymbol{z}$ is sampled from a categorial, or multinomial, distribution with the mass probability $\boldsymbol{\pi} = \left[\pi_1, \pi_2, \ldots, \pi_n\right]^T$ such that $\pi_i \geq 0$, $\sum_{i=1}^{n} \pi_i = 1$ , i.e. $p\left(z = k\right) = \pi_k$.

$$\begin{aligned} \boldsymbol{z},\boldsymbol{\theta} \quad &\sim \quad \text{Categorial}\left(\boldsymbol{\pi}\right) \\ \boldsymbol{x}|\,\boldsymbol{z},\boldsymbol{\theta} \quad &\sim \quad \textstyle\sum_{k=1}^{K} \pi_k \mathcal{N}\left(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right) \end{aligned} \qquad (3)$$

In this case, $\boldsymbol{\theta}$ are the set of unknown parameters such as $\pi_1$, $\boldsymbol{\mu}_1$,$\boldsymbol{\Sigma}_1$, etc. Given a set of data $\{\boldsymbol{x}_i\}$ , $i = 1, 2, ..., N$, the model log-likelihood is

$$\log p\left(X|\,\boldsymbol{\theta}\right) = \sum_{i=1}^{N} \log \left[\sum_{k=1}^{K} \pi_i p\left(\boldsymbol{x}_i|\,\boldsymbol{z}_k\right)\right] \qquad (4)$$

The update formulas for $k^{th}$ Gaussian component at iteration $j$ are given by [5]:

$$\left\langle z_{k,i}^{(j)} \right\rangle = \frac{\pi_k \mathcal{N}\left(\boldsymbol{x}_i|\,\boldsymbol{\mu}_k^{(j)}, \boldsymbol{\Sigma}_k^{(j)}\right)}{\sum_{m=1}^{M} \pi_m \mathcal{N}\left(\boldsymbol{x}_i|\,\boldsymbol{\mu}_m^{(j)}, \boldsymbol{\Sigma}_m^{(j)}\right)} \qquad (5a)$$

$$\pi_k^{(j+1)} = \frac{1}{N} \sum_{i=1}^{N} \left\langle z_{k,i}^{(j)} \right\rangle \qquad (5b)$$

$$\boldsymbol{\mu}_k^{(j+1)} = \frac{\sum_{i=1}^{N} \left\langle z_{k,i}^{(j)} \right\rangle \boldsymbol{x}_i}{\sum_{i=1}^{N} \left\langle z_{k,i}^{(j)} \right\rangle} \qquad (5c)$$

$$\boldsymbol{\Sigma}_k^{(j+1)} = \frac{1}{\sum_{i=1}^{N} \left\langle z_{k,i}^{(j)} \right\rangle} \sum_{i=1}^{N} \left\langle z_{k,i}^{(j)} \right\rangle \left(\boldsymbol{x}_i - \boldsymbol{\mu}_k^{(j)}\right) \left(\boldsymbol{x}_i - \boldsymbol{\mu}_k^{(j)}\right)^T \qquad (5d)$$

In variational inference view, let $q\left(\boldsymbol{h}, \boldsymbol{\phi}\right)$ be the probability distribution on some hidden variables $\boldsymbol{h} \in \mathbb{R}^d$ with some parameters $\boldsymbol{\phi}$, for simplicity, the $\phi$-dependency is implicitly acknowledged, $\phi$ will be dropped from the notation. Since $\int q\left(\boldsymbol{z}\right) \mathrm{d}\boldsymbol{z} = 1$, the log-likelihood can be re-written as [6]

$$\log p\left(\boldsymbol{x}|\,\boldsymbol{\theta}\right) = \mathcal{L}\left(q, \boldsymbol{\theta}\right) + KL\left(q \parallel p_{\boldsymbol{h}|\boldsymbol{x}}\right) \qquad (6)$$

with

$$\mathcal{L}\left(q, \theta\right) = \int q\left(\boldsymbol{h}\right) \log \left(\frac{p\left(\boldsymbol{x}, \boldsymbol{h}|\,\theta\right)}{q\left(\boldsymbol{h}\right)}\right) \mathrm{d}\boldsymbol{h} \qquad (6a)$$

and

$$KL\left(q \parallel p_{\boldsymbol{h}|\boldsymbol{x}}\right) = -\int q\left(\boldsymbol{h}\right) \log \left(\frac{p\left(\boldsymbol{h}|\,\boldsymbol{x}, \theta\right)}{q\left(\boldsymbol{h}\right)}\right) \mathrm{d}\boldsymbol{h} \qquad (6b)$$

The term $KL\left(q \parallel p\right) \geq 0$, $\forall p, q$ is known as the Kullback-Leibler divergence which measures how much $q$ differs from $p$. Hence, (6) becomes

$$\log p\left(\boldsymbol{x}|\,\boldsymbol{\theta}\right) \geq \mathcal{L}\left(q, \boldsymbol{\theta}\right) \qquad (7)$$

$\mathcal{L}\left(q, \boldsymbol{\theta}\right)$ is called the evidence lower bound (ELBO). As can be seen from (7), the ELBO provides a lower bound to the likelihood, which makes maximizing it is as good as maximizing the likelihood. The gap (difference) between the log-likelihood and the ELBO is exactly the KL divergence, if $KL\left(q \parallel p\right) = 0$, the ELBO hits the log-likelihood, $q$ is identical to $p_{\boldsymbol{h}|\boldsymbol{x}}$. Variational inference focuses on choosing a suitable variational distribution, i.e. the form of $q$, that best approximates the true distribution $p\left(\boldsymbol{h}|\,\boldsymbol{x}, \theta\right)$ by optimizing the ELBO through coordinate ascent [7].

One main focus of variational inference approach is to choose the form of $q$. In this work, we use a mean-field approximation [5] to $q\left(\boldsymbol{h}\right)$, i.e. the variational variables are factorizable, each variational distribution $q_i$ is choosen from the exponential family

$$q(\boldsymbol{h}) = \prod_{t=1}^{d} q_i(h_i) \tag{8}$$

The ELBO can be written as [7]

$$\mathcal{L}(q, \boldsymbol{\theta}) \sim \sum_{t=1}^{d} \mathbb{E}_q \left[\log p\left( h_t \middle| h_{\bar{t}}, X\right)\right] - \mathbb{E}_{q_t}\left[\log q_t\left(h_t\right)\right] \tag{9}$$

where the subscript $\bar{t}$ means all indices other than $t$. Setting the derivative of $\mathcal{L}(q, \theta)$ to 0, we arrive at coordinate ascent update

$$q^*(h_t) \sim \mathbb{E}_{q_{\bar{t}}}\left[\log p\left( h_t \middle| h_{\bar{t}}, X\right)\right] \tag{10}$$

Once a conjugate prior is put over the parameters, the update rule in (10) is tractable and analytical.

## IV. Examples

In order to test the validity of our method, two types of total jitter construction are performed: one consists of PJ and RJ, and another has DDJ and RJ. Since the pdf of a typical, single frequency PJ has highest density at its bounds, it is reasonable to approximate the pdf of PJ as dual-dirac. As shown in Table I, there is a total of 7 jitter combinations based on the selection of RJ and PJ/DDJ. The term PJ/DDJ in Table 1 is used to describe $|\mu_L$ - $\mu_R|$ while the term RJ tells the size of $\sigma$. The amount of 10,000 total jitter samples are collected for extraction. The jitter estimated value and jitter error are recorded.

TABLE I
Jitter Study Cases

| Cases | Jitter Added (ps) |
|---|---|
| Case 1 | RJ = 10, PJ = 5 |
| Case 2 | RJ = 15, PJ = 5 |
| Case 3 | RJ = 5, PJ/DDJ = 5 |
| Case 4 | RJ = 10, PJ/DDJ = 10 |
| Case 5 | RJ = 15, PJ/DDJ = 15 |
| Case 6 | RJ = 5, DDJ = 10 |
| Case 7 | RJ = 5, DDJ = 15 |

Comparing the DJ error in Figure 3, SVI shows less error in both PJ and DDJ cases than EM. As expected, in cases that have PJ, when approximating the PJ as dual-delta, DJ estimation error is larger than that in cases that have DDJ. It is worth noting that this approximation gets worse (over 100% error) when RJ is the dominant contributor in the underlying jitter. The reason could be the distinction between the dual-RJ in total jitter (TJ) is ambiguous but SVI is still advanced by the half error rate of EM method.

In Figure 4, SVI method performs significantly better than the EM method when estimating RJ.

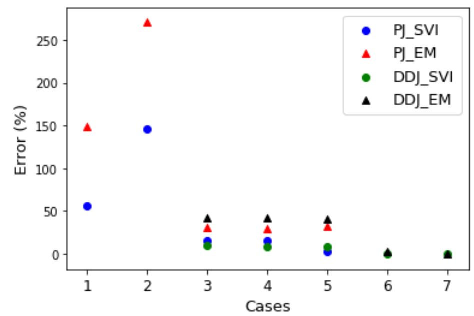


Fig. 3. DJ error between true and extracted values

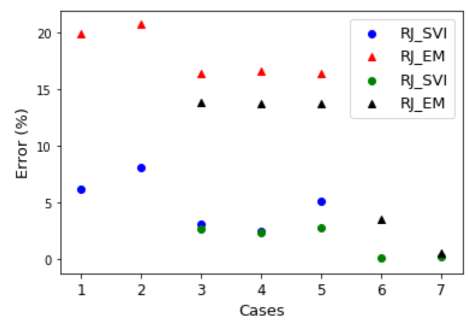


Fig. 4. RJ error between true and extracted values

## V. Conclusion

An efficient and accurate jitter decomposition method that separates periodic jitter or data-dependent jitter and random jitter is presented. This method is based on stochastic variational inference of Gaussian mixture model to update and estimate jitter parameters. The comparison between this novel method and traditional GMM demonstrates the accuracy of the proposed jitter separation method. In future work, we will demonstrate the robustness of SVI in compiling more complicated graphical models to extract more complicated jitter mixtures.

## References

[1] M. P. Li, J. Wilstrup, R. Jessen, and D. Petrich, "A new method for jitter decomposition through its distribution tail fitting," in *International Test Conference 1999. Proceedings (IEEE Cat. No.99CH37034)*, pp. 788–794, 1999.

[2] F. Nan, Y. Wang, F. Li, W. Yang, and X. Ma, "A better method than tail-fitting algorithm for jitter separation based on gaussian mixture model," *J. Electronic Testing*, vol. 25, pp. 337–342, Dec 2009.

[3] D. Mistry, S. Joshi, and N. Agrawal, "A novel jitter separation method based on gaussian mixture model," in *2015 International Conference on Pervasive Computing (ICPC)*, pp. 1–4, 2015.

[4] R. Stephens, "Jitter analysis: The dual-dirac model, RJ/DJ, and Qscale." White Paper, Agilent Technologies, Dec 2004.

[5] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.

[6] R. Neal and G. Hinton, "A view of the EM algorithm that justifies incremental, sparse, and other variants," in *Learning in Graphical Models* (M. I. Jordan, ed.), pp. 355–368, MIT Press, 1999.

[7] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *Journal of the American Statistical Association*, vol. 112, pp. 859–877, Feb 2017.