Joint compensation of jitter noise and time shift errors in multichannel sampling system

Hesam Araghi, Mohammad Ali Akhaee, *Member, IEEE,* and Arash Amini, *Senior Member, IEEE*

***Abstract*—In high-speed analog to digital converters (ADCs), two main factors contribute to high power consumption. The first is the super-linear relationship with the sampling rate; *i.e.* by doubling the sampling rate, the power consumption more than doubles. The second factor arises from the consumption of analog circuitry responsible to mitigate the jitter noise. By employing a multichannel sampling system, one can achieve high sampling rates by incorporating multiple low sampling-rate channels, which results in linear scaling of power consumption with the number of channels. The main drawback of this system is the timing mismatch between the sampling channels. In this paper, we intend to jointly compensate the jitter noise and the timing mismatch between the channels using statistical methods. We first approximate the acquisition system and derive a stochastic model. Then, we propose an iterative maximum likelihood (ML) algorithm to estimate the parameters of the input signal. We further evaluate the Cramér-Rao lower bound (CRLB) for the estimation error to examine the proposed algorithm. Simulation results indicate that our algorithm is capable of closely following the CRLB curve for reasonable values of jitter noise and wide range of timing mismatch errors. Moreover, it is shown that the mismatch-compensated multichannel sampling system performs almost equivalently to a single-channel high rate sampler without having its shortcomings.**

***Index Terms*—Cramér-Rao lower bound, jitter noise, maximum likelihood estimation, multichannel sampling system, timing mismatch compensation.**

I. Introduction

VAST and efficient processing capabilities of high-speed microprocessors have led to a major shift in signal processing from continuous to digital domain within the past decades. Nowadays, it is difficult to find a piece of technology that does not benefit from digital data processor units. In devices that take continuous-domain signals as input, the signal needs to be discretized (digitized) before being processed by digital processors. The block responsible for the discretization task is called an analog-to-digital converter (ADC). An ideal ADC uniformly samples the input signal according to a given sampling rate.

The practical limitations on the sampling rates of current ADCs is a major challenge in modern applications that deal with wideband signals. Examples of the applications with high sampling rate requirement include software-defined radio receivers [1], [2], ultra-wideband communications [3], [4], and spectrum sensing for cognitive radio networks [5], [6]. However, ADCs with high conversion rates are also power-consuming and very costly [7].

One simple technique to tackle the power and cost problems is to utilize multiple low-rate ADCs in a time-interleaved (TI) structure instead of one high-speed ADC [8]. A TI-ADC of sampling rate $C\,f_s$ passes the analog signal into C identical parallel ADC channels where each channel samples the signal at the rate f_s. Ideally, the timing clock of the ith ADC has a delay of $i/(C\,f_s)$ relative to the reference channel $i = 0$. The output of the TI-ADC can be formed by multiplexing the channel samples into a single high-rate output; nevertheless, it is sometimes easier to separately study the channels (including our approach in this paper). Although the single output of a TI-ADC is essentially the same as a high-rate ADC, it does suffer from power dissipation as much [9].

While parallel usage of low-rate and low-power ADCs in form of a TI-ADC helps in increasing the overall sampling rate and power efficiency (and decreasing the cost), the synchronization of channels could potentially be a challenge. Further, varying specifications of the ADCs in different channels, which is known as manufacturing mismatch, degrades the performance. Due to limitations in physical layout, variations in the signal and clock path lengths, and different gate delays of the channels, such mistmatches are unavoidable [8], [10]–[12].

The three major types of mismatches in the time interleaved structure are offset, gain, and timing mismatches. Their adversarial effects in the sampling process have been extensively studied in [13]–[15].

- **Gain mismatch:** The amplifiers of the ADCs are not exactly identical. Hence, the input signals are subject to different gains (g_i).
- **Offset mismatch:** When the ADCs do not have a unified voltage ground, they might differ by additive amplitude mismatch constants (a_i).
- **Timing mismatch:** The requirement of equal time delays between consecutive sampling channels in a TI-ADC may not be accurately satisfied. As a result, the ADCs sample the input signal with a different but fixed time delays (t_i) from their nominal sampling instances. This mismatch is known as the timing mismatch and results in a periodic nonuniform sampling structure.

Among these three types of mismatches, compensating the timing mismatch is more challenging due to its non-linear, signal-dependent nature [15]. In this paper, we focus on estimating and compensating the timing mismatch in a TI-ADC by ignoring the two other sources of mismatch, which are straightforward to correct. The compensation of offset and gain mismatches have been previously studied in [16]–[18].

Another source of timing error in general ADCs (not limited to TI-ADCs) is the jitter noise. This noise disrupts the regularity of the clock signal and results in unpredictable deviation of the sampling instances from the nominal equidistant structure

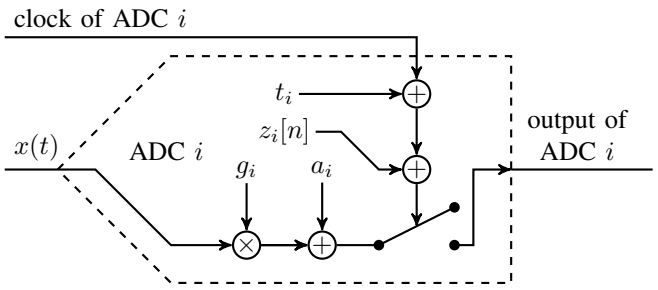


Fig. 1. The mathematical model of the ith channel in a TI-ADC. The input signal $x(t)$ experiences gain (scaling with g_i) and offset mismatches (addition with a_i). The constant offset of the clock signal is modeled via t_i, while the random and sample-dependent jitter noise is shown in the form of $z_i[n]$.

[19], [20]. This effect could be reduced in the analog domain by employing clock sources with relatively low phase noise [20], [21]. Unfortunately, the required circuits for this purpose impose high power-consumption [22]. The other alternative is to apply signal processing techniques in the digital domain (after the samples are taken) to compensate for the timing errors [23]. Due to the advancement of microelectronic hardware, this solution is quite low-power and low-cost.

Fig. 1 depicts the block diagram for the equivalent model of all the discussed mismatches for a single channel of a TI-ADC. In this figure, besides the desired time shift of $i/(C\,f_s)$ in the ith channel, the sampling clock is perturbed by a constant but unwanted time shift t_i, as well as the stochastic jitter noise $z_i[n]$ which independently varies over time. The amplitude and gain mismatches are also modeled by a_i and g_i in this figure, respectively. As explained earlier, a_i is applied in an additive manner, while g_i has a multiplicative form.

A. Related work

Compensation of timing mismatch in multichannel sampling structures has been widely studied during past few years [11], [13], [15], [24]–[32]. The methods can be divided into two main categories based on the available knowledge about the input signal. The category of "foreground" techniques refer to the cases where a known signal is injected into the TI-ADC for the estimation of the offsets [27], [31], [32]. By having the estimates, there exist various approaches for the compensation, namely, least squares approach [26], time-varying filtering [28], and Lagrange interpolation [31], [32]. While foreground techniques usually result in more accurate estimations, they interrupt the normal operation of the TI-ADC. To avoid this drawback, one might opt for "background" (or blind) techniques, which only require the knowledge about the model of the input (e.g., correlation pattern, or bandlimitedness), instead of the actual input. These methods estimate the offsets gradually and adaptively. A more elaborate technique in [25] proposes a rank minimization problem instead. Since the variables in the minimization techniques are the set of channel time shifts, the size of the problem grows as the number of channels increases. Hence, the computational complexity of the estimation method increases. In practice, the computational cost is tolerable only for TI-ADCs with small number of channels. By considering the out-of-band energy of the signal caused by aliasing, a least mean-square method is presented in [29]. A connection between the I/Q mismatch problem and the timing offset compensation is exploited in [30] for a four-channel TI-ADC. The method works by identifying and suppressing spurious frequency components in the sampled signal.

Digital mitigation of jitter noise in the ADCs dates back to the work of Balakrishnan [33]. In this work, the effect of jitter noise is studied for a bandlimited signal with both deterministic and stochastic approaches. Also, optimal basis functions are derived that result in minimum mean–square error among linear methods. In [34], probability density function (pdf) of the jitter noise is approximated by a parametric model. The author utilizes a minimization algorithm to determine the parameters. Then, the compensated waveform is obtained with a pdf deconvolution method. Authors in [35] present a maximum likelihood estimator to determine the additive and the jitter noise variances. For the maximization part, they employ a Newton–like method and compare their results with the least squares solution. The optimal linear estimator for the amplitude noise in presence of jitter noise is evaluated in [36]. The mean-square-error study of reconstrcuting periodic bandlimited signals in [37] reveals that the error decays as $\mathcal{O}(1/n)$, where n stands for the number of samples per cycle. Weller and Goyal establish a Bayesian framework for the jitter mitigation problem and develop a non-linear estimator based on Gibbs sampling method [38]. They also implement an Expectation-Maximization (EM) algorithm using numerical integration to approximate the ML estimation of input signal parameters [39]. However, Gibbs sampling and EM methods are computationally expensive. A low-cost approximate ML estimator is proposed in [40], that assumes jitter noise values are small and employs truncation of the corresponding Taylor series.

B. Contribution

In this paper, our goal is to *jointly compensate the effects of both jitter noise and time shift errors* in a time-interleaved sampling structure. The obtained results show that we can use low-power clock sources with higher jitter noise to achieve similar performances. In [41], a numerical approach of the Bayesian model for the joint compensation problem is introduced. The iterative ML algorithm for the simpler approximation model of the TI-ADC structure are presented in [42], where it can be viewed as an extension of [40] with non-zero-mean location errors. In [42], it is assumed that the sum of jitter noise and time shift errors are Gaussian random variables and can be approximated by Taylor expansion. While this assumption can be true for jitter noise, time shifts errors may have larger values which violate the assumption. Hence in the current paper, we do not assume small time shifts errors. We first derive an approximate acquisition model for the problem with deterministic unknown time shift errors. Then, we propose a new block-based iterative ML algorithm which eliminates the need to assume small values for time shift errors. The algorithm is capable of both estimating and compensating these large values. After collecting the

necessary number of samples from each different channels, this block-based postprocessing algorithm can be performed to estimate the signal parameters either by an off-chip computer or an on-chip processing unit installed after the acquisition system. Moreover, the Cramér-Rao lower bound (CRLB) of the approximate model is computed and used as a performance quality measure.

C. Notation

Throughout this paper, vectors are denoted by boldface lowercase letters, e.g. $\mathbf{x}$ and $\boldsymbol{\eta}$, and matrices by boldface uppercase letters, e.g. $\mathbf{A}$ and $\boldsymbol{\Sigma}$. We use $(\cdot)^T$ for transpose of a vector or matrix and brackets for indexing, e.g. $\mathbf{x}[k]$ or $\mathbf{A}[m,n]$. In addition, the ith column of the matrix $\mathbf{A}$ is shown by $\mathbf{A}[:,i]$. We use $p(\mathbf{x})$ notation for probability density function of $\mathbf{x}$ and $p(\mathbf{x};\mathbf{y})$ for conditional pdf $\mathbf{x}$ given nonrandom parameter vector $\mathbf{y}$. $\mathcal{N}(\eta,\sigma^2)$ represents the pdf for the special case of Gaussian variable with mean η and variance σ^2; similarly, $\mathcal{N}(\boldsymbol{\eta},\boldsymbol{\Sigma})$ is the multivariate Gaussian distribution with mean vector $\boldsymbol{\eta}$ and covariance matrix $\boldsymbol{\Sigma}$. We reserve the notations $\mathbb{E}[\cdot]$ for the operator of statistical expectation. Further, $\mathrm{tr}\{\cdot\}$ stands for the trace operator for square matrices. $\mathbf{A} \succcurlyeq \mathbf{B}$ implies that the matrix $\mathbf{A}-\mathbf{B}$ is positive semidefinite. Finally, $\mathrm{diag}(\mathbf{x})$ is a diagonal matrix whose diagonal entries correspond to the elements in the vector $\mathbf{x}$:

$$\mathrm{diag}(\mathbf{x})[i,j] = \begin{cases} x_i & i=j \\ 0 & i\neq j \end{cases}$$

Similarly, block-diagonal matrix operator is defined as

$$\mathrm{blkdiag}(\mathbf{A}_0,\mathbf{A}_1,\ldots,\mathbf{A}_{N-1}) = \begin{bmatrix} \mathbf{A}_0 & & & 0 \\ & \mathbf{A}_1 & & \\ & & \ddots & \\ 0 & & & \mathbf{A}_{N-1} \end{bmatrix}.$$

D. Outline

The rest of this paper is organized as follows. In Section II, we describe the acquisition model and incorporate the jitter noise and timing mismatch effects. Section III presents the iterative ML method for the estimation of the unknown parameters of the input signal and the time shift errors. We derive the Cramér-Rao lower bound for the MSE of the approximate model in Section IV. Numerical simulations in Section V show the performance of the proposed method and demonstrates the CRLB results of Section IV. Discussion of computational complexity and sub-Nyquist sampling scenario, as well as practical remarks can be found in Section VI. Finally, Section VII concludes the paper.

II. Acquisition Model

In this section, we describe mathematical models for the TI-ADC structure as well as the input signal in separate subsections.

A. TI-ADC structure

To model the TI-ADC structure, we suppose that there exist C parallel channels and that each of them samples the input signal at an oversampling ratio of M where $M \geq 1$, resulting in an overall MC oversampling ratio. The ideal set of sampling instances for the ith channel is given by

$$\mathcal{T}_i = \left\{ n\,T_s + i\tfrac{T_s}{C} \right\}_{n\in\mathbb{Z}}, \tag{1}$$

where $T_s = 1/f_s$ is the sampling period in each channel, and $0 \leq i \leq C-1$. However, according to Fig. 1, we perturb the sampling instances in each channel by an additive constant t_i and time varying values $z_i[n]$ to model the time shift error and jitter noise, respectively. Hence, the obtained model for the sampling instances shall be

$$\widetilde{\mathcal{T}}_i = \left\{ n\,T_s + i\tfrac{T_s}{C} + z_i[n] + t_i \right\}_{n\in\mathbb{Z}}. \tag{2}$$

By including the amplitude noise, the output of the ith channel is then formulated as

$$y_i[n] = x\big(n\,T_s + i\tfrac{T_s}{C} + z_i[n] + t_i\big) + w_i[n]. \tag{3}$$

Here, we assume that $w_i[n]$s are independent and identically distributed (i.i.d.) zero-mean Gaussian variables with variance σ_w^2 to model the joint effect of thermal and quantization noises. We also use zero-mean Gaussian noise with variance σ_z^2 for $z_i[n]$s which are independent of $w_i[n]$s. To set a basis for the time shifts, we assume $t_0 = 0$.

B. Input signal model

The input signal $x(t)$ is assumed to lie in the subspace of Hilbert space $L^2(\mathbb{R})$ with basis $\{h_k(t) = h(t/T-k) : k\in\mathbb{Z}\}$:

$$x(t) = \sum_{k\in\mathbb{Z}} x(k)\,h\big(\frac{t}{T}-k\big), \tag{4}$$

where $\{x(k)\}_{k\in\mathbb{Z}}$ is the set of deterministic signal parameters. One important example of $h(t)$ used throughout this paper is $\mathrm{sinc}(t) = \sin(\pi t)/\pi t$ which by choosing $T = T_c$ corresponds to bandlimited signal subspace with maximum frequency of $f_c/2 = 1/2T_c$. If we choose $f_c = 1$ (without loss of generality), the famous Shannon sampling theorem states that $x(t)$ can be written as

$$x(t) = \sum_{k\in\mathbb{Z}} x(k)\ \mathrm{sinc}\big(t-k\big). \tag{5}$$

where $x(k)$ denotes the uniform samples at the Nyquist rate of $f_c = 1$, and $\mathrm{sinc}(t) = \sin(\pi t)/\pi t$. By evaluating (5) at the actual sampling instances provided in (2) and replacing T_s with $\frac{1}{M\,f_c} = \frac{1}{M}$ to take the oversampling ratio into account, we can rewrite (3) as

$$y_i[n] = \sum_{k\in\mathbb{Z}} x(k)\ \mathrm{sinc}\big(\tfrac{n}{M} + \tfrac{i}{MC} + z_i[n] + t_i - k\big) + w_i[n]. \tag{6}$$

The recovery of the original signal $x(t)$ is now equivalent to obtaining the samples $x(k)$ based on the observations $\{y_i[n]\}_i$. The infinite summation in (6) relates each observed sample to an infinite set of noiseless samples $x(k)$ of the original

signal. In practice, however, the observation period is always limited in time (finite number of observations). Moreover, the estimation of the signal beyond the observed time window is rarely a priority. Hence, we approximate (6) as

$$y_i[n] \approx \sum_{k=0}^{K-1} x(k)\,\mathrm{sinc}\left(\tfrac{n}{M} + \tfrac{i}{M\,C} + z_i[n] + t_i - k\right) + w_i[n], \quad (7)$$

where the observation period is limited to $t \in [0, K)$ and the set of observations in the ith channel shall be $\{y_i[n]\}_{n=0}^{N_C-1}$ with $N_C = M\,K$. In equation (7), the terms beyond $0 \le k \le K-1$ are ignored in the summation which can add some error to the acquired samples especially at the borders. To reduce the effect of the truncation error, one may prefer to use a fast decaying function for $h(t)$ instead of sinc function with slow decay, e.g. raised cosine or wavelet scaling functions. With this approximation, the recovery problem translates into estimating $\{x(k)\}_{k=0}^{K-1}$ based on the available $C\,N_C$ observations of a C-channel TI-ADC. Equation (7) can be expressed in a matrix form as

$$\mathbf{y}_i = \mathbf{H}_i(t_i, \mathbf{z}_i)\mathbf{x} + \mathbf{w}_i, \quad (8)$$

where

$$\begin{aligned}
\mathbf{y}_i &= [y_i[0], \ldots, y_i[N_C-1]]^T,\\
\mathbf{x} &= [x[0], \ldots, x[K-1]]^T,\\
\mathbf{z}_i &= [z_i[0], \ldots, z_i[N_C-1]]^T,\\
\mathbf{w}_i &= [w_i[0], \ldots, w_i[N_C-1]]^T,\\
\mathbf{H}_i(t_i, \mathbf{z}_i)[n,k] &= \mathrm{sinc}(\tfrac{n}{M} + \tfrac{i}{M\,C} + z_i[n] + t_i - k).
\end{aligned}$$

Here, the variables t_i and $\mathbf{z}_i$ in the notation $\mathbf{H}(t_i, \mathbf{z}_i)$ highlight the fact that this matrix depends on these variables.

III. Iterative ML estimator

In this section, we develop an iterative Maximum Likelihood (ML) method to estimate the signal parameters which are the Nyquist-rate samples of the input signal in the observation period.

In practice, the Jitter noise is relatively small compared to the sampling period $1/(M\,C)$. Based on this fact, we eliminate the non-linear effect of jitter noise ($\mathbf{z}_i$) by applying the first-order Taylor series approximation to (7):

$$\begin{aligned}
y_i[n] \approx &\sum_{k=0}^{K-1} x(k)\,\mathrm{sinc}\left(\tfrac{n}{M} + \tfrac{i}{M\,C} + t_i - k\right)\\
&+ \sum_{k=0}^{K-1} x(k)\,\mathrm{sinc}^{(1)}\left(\tfrac{n}{M} + \tfrac{i}{M\,C} + t_i - k\right) z_i[n] + w_i[n], \quad (9)
\end{aligned}$$

where $\mathrm{sinc}^{(1)}(x) = \frac{\mathrm{d}}{\mathrm{d}x}\mathrm{sinc}(x)$. The matrix form of (9) is given by

$$\mathbf{y}_i \approx \mathbf{H}_i(t_i, \mathbf{0})\,\mathbf{x} + \mathbf{Z}_i\,\mathbf{D}_i(t_i)\,\mathbf{x} + \mathbf{w}_i, \quad (10)$$

where $\mathbf{Z}_i = \mathrm{diag}(\mathbf{z}_i)$ and $\mathbf{D}_i(t_i)$ is a $N_C \times K$ matrix with the elements

$$\mathbf{D}_i(t_i)[n,k] = \mathrm{sinc}^{(1)}\left(\tfrac{n}{M} + \tfrac{i}{M\,C} + t_i - k\right). \quad (11)$$

In our approach, we treat the signal parameters $\mathbf{x}$ and the time shift errors of the channels t_is as deterministic but unknown variables of the problem. This helps us conclude that given $\mathbf{x}$ and t_is, $\mathbf{y}_i$s are jointly normal random vectors. Indeed, each $y_i[n]$ can be seen as a linear combination of independent Gaussian random variables $z_i[n] \sim \mathcal{N}(0, \sigma_z^2)$ and $w_i[n] \sim \mathcal{N}(0, \sigma_w^2)$. Therefore, $\mathbf{y}_i$ is a Gaussian random vector

$$\mathbf{y}_i \sim \mathcal{N}(\boldsymbol{\eta}_i, \boldsymbol{\Sigma}_{\mathbf{y}_i}), \quad (12)$$

with mean

$$\boldsymbol{\eta}_i = \mathbb{E}[\mathbf{y}_i] = \mathbf{H}_i(t_i, \mathbf{0})\,\mathbf{x}, \quad (13)$$

and covariance

$$\begin{aligned}
\boldsymbol{\Sigma}_{\mathbf{y}_i} &= \mathbb{E}\left[\left(\mathbf{y}_i - \mathbf{H}_i(t_i,\mathbf{0})\,\mathbf{x}\right)\left(\mathbf{y}_i - \mathbf{H}_i(t_i,\mathbf{0})\,\mathbf{x}\right)^T\right]\\
&= \mathbb{E}\left[\mathbf{Z}_i\left(\mathbf{D}_i(t_i)\,\mathbf{x}\,\mathbf{x}^T\,\mathbf{D}_i(t_i)^T\right)\mathbf{Z}_i\right] + \sigma_w^2\,\mathbf{I}_{N_C}\\
&= \mathbb{E}\left[\mathbf{z}_i\,\mathbf{z}_i^T\right] \odot \left(\mathbf{D}_i(t_i)\,\mathbf{x}\,\mathbf{x}^T\,\mathbf{D}_i(t_i)^T\right) + \sigma_w^2\,\mathbf{I}_{N_C} \quad (14a)\\
&= \boldsymbol{\Sigma}_{\mathbf{z}_i} \odot \left(\mathbf{D}_i(t_i)\,\mathbf{x}\,\mathbf{x}^T\,\mathbf{D}_i(t_i)^T\right) + \sigma_w^2\,\mathbf{I}_{N_C}, \quad (14b)
\end{aligned}$$

where $\odot$ denotes the Hadamard matrix product and the first term in (14a) is obtained by recalling that the matrix $\mathbf{Z}_i$ is diagonal. Here, $\boldsymbol{\Sigma}_{\mathbf{z}_i}$ is the covariance matrix of jitter noise $\mathbf{z}_i$. In general, equation (14b) is used for the correlated jitter noise. However, $\boldsymbol{\Sigma}_{\mathbf{y}_i}$ can further simplify to a diagonal matrix for the uncorrelated case as below

$$\boldsymbol{\Sigma}_{\mathbf{y}_i} = \sigma_z^2\,\mathbf{I}_{N_C} \odot \left(\mathbf{D}_i(t_i)\,\mathbf{x}\,\mathbf{x}^T\,\mathbf{D}_i(t_i)^T\right) + \sigma_w^2\,\mathbf{I}_{N_C}, \quad (15)$$

where σ_z^2 is the variance of jitter noise. We assume uncorrelated jitter noise throughout the paper, but our proposed method can be extended easily to the correlated case by using equation (14b) instead of (15).

For the random vectors $\mathbf{w}_i$ and $\mathbf{z}_i$ with $0 \le i, j \le C-1$ and $i \ne j$ we have that

$$\mathbf{z}_i \perp\!\!\!\perp \mathbf{w}_i \qquad \mathbf{z}_i \perp\!\!\!\perp \mathbf{z}_j \qquad \mathbf{w}_i \perp\!\!\!\perp \mathbf{w}_j$$

Here, $\mathbf{a} \perp\!\!\!\perp \mathbf{b}$ denotes the statistical independence of vectors $\mathbf{a}$ and $\mathbf{b}$. Thus, given $\mathbf{x}$ and t_is, the vectors $\mathbf{y}_i$ are pairwise independent:

$$\mathbf{y}_i \perp\!\!\!\perp \mathbf{y}_j \;\; (i \ne j). \quad (16)$$

From (16), the pdf of $\mathbf{y} = [\mathbf{y}_0^T, \mathbf{y}_1^T, \ldots, \mathbf{y}_{C-1}^T]^T$ can be written as

$$p(\mathbf{y}; \mathbf{x}, \mathbf{t}) = \prod_{i=0}^{C-1} p(\mathbf{y}_i; \mathbf{x}, t_i). \quad (17)$$

Hence, the log likelihood function of $\mathbf{x}$, $\{t_i\}_i$ after observing $\breve{\mathbf{y}}$ is

$$\begin{aligned}
LL(\mathbf{x}, \mathbf{t}; \breve{\mathbf{y}}) &\triangleq \log p(\breve{\mathbf{y}}; \mathbf{x}, \mathbf{t}) = \sum_{i=0}^{C-1} \log p(\breve{\mathbf{y}}_i; \mathbf{x}, \mathbf{t})\\
&= -\frac{CN_C}{2}\log(2\pi) - \frac{1}{2}\sum_{i=0}^{C-1}\log(\det(\boldsymbol{\Sigma}_{\mathbf{y}_i}))\\
&\quad - \frac{1}{2}\sum_{i=0}^{C-1}(\breve{\mathbf{y}}_i - \mathbf{H}_i(t_i,\mathbf{0})\,\mathbf{x})^T\boldsymbol{\Sigma}_{\mathbf{y}_i}^{-1}(\breve{\mathbf{y}}_i - \mathbf{H}_i(t_i,\mathbf{0})\,\mathbf{x}), \quad (18)
\end{aligned}$$

where $\breve{\mathbf{y}}$ is the observing vector of all channels output and $\breve{\mathbf{y}}_i$ is the observing vector of ith channel output.

The Maximum Likelihood (ML) estimate of $\mathbf{x}$ given $\breve{\mathbf{y}}$ is found based on the maximizer of the LL in (18) as a function of $\mathbf{x}$ and $\mathbf{t}$.

$$\big(\hat{\mathbf{x}}_{ML}, \hat{\mathbf{t}}_{ML}\big) = \arg\max_{\mathbf{x},\mathbf{t}} LL(\mathbf{x}, \mathbf{t}; \breve{\mathbf{y}}). \tag{19}$$

As it can be seen in (15), the covariance matrix $\mathbf{\Sigma}_{\mathbf{y}_i}$ depends on both $\mathbf{x}$ and t_i. Therefore, it is not straightforward to find a closed form expression for the ML estimator. In order to tackle the problem, we use an iterative method to approximately maximize the LL function. First, we fix t_is and maximize the function with respect to $\mathbf{x}$; next, we fix $\mathbf{x}$ by the updated values and maximize the function with respect to t_is. Before each step, we evaluate $\mathbf{\Sigma}_{\mathbf{y}_i}$ based on the available $\mathbf{x}$ and t_is; however, we ignore the dependency of the covariance matrix on both $\mathbf{x}$ and t_is in each maximization. As indicated in (15), the variations of the covariance matrix are negligible for small jitter noise values ($\sigma_z \ll 1$).

A. Maximization with respect to $\mathbf{x}$

By fixing $t_i^{(j)}$s (jth iteration) and the covariance matrices $\mathbf{\Sigma}_{\mathbf{y}_i}$s, the LL function becomes quadratic in terms of $\mathbf{x}$ and its gradient is given by

$$\frac{\partial}{\partial \mathbf{x}} LL\big(\mathbf{x}, \mathbf{t}^{(j)}; \breve{\mathbf{y}}\big) = \sum_{i=0}^{C-1} \mathbf{H}_i\big(t_i^{(j)}, \mathbf{0}\big)^T \mathbf{\Sigma}_{\mathbf{y}_i}^{-1} \Big(\breve{\mathbf{y}}_i - \mathbf{H}_i\big(t_i^{(j)}, \mathbf{0}\big)\, \mathbf{x}\Big). \tag{20}$$

Hence, the maximizer for which the gradient vanishes could be written as

$$\mathbf{x}^{(j+1)} = \Big[\sum_{i=0}^{C-1} \mathbf{H}_i\big(t_i^{(j)}, \mathbf{0}\big)^T \mathbf{\Sigma}_{\mathbf{y}_i}^{-1} \mathbf{H}_i\big(t_i^{(j)}, \mathbf{0}\big)\Big]^{-1} \times \sum_{i=0}^{C-1} \mathbf{H}_i\big(t_i^{(j)}, \mathbf{0}\big)^T \mathbf{\Sigma}_{\mathbf{y}_i}^{-1} \breve{\mathbf{y}}_i. \tag{21}$$

As it can be seen $\mathbf{x}^{(j+1)}$ is a function of $t_i^{(j)}$s; further, it depends on $\mathbf{x}^{(j)}$, σ_z^2, and σ_w^2 through $\mathbf{\Sigma}_{\mathbf{y}_i}^{-1}$. Here, we assume the values of σ_z^2 and σ_w^2 are available. For example, we can estimate these variances using the method introduced in [35]. In Subsection V-B, we investigate the sensitivity of the proposed algorithm to the error in these values.

B. Maximization with respect to $\mathbf{t}$

Similar to the previous step, we fix $\mathbf{x}^{(j+1)}$ and the covariance matrices $\mathbf{\Sigma}_{\mathbf{y}_i}$s, and maximize the LL function in terms of $\mathbf{t}$. We have

$$\frac{\partial}{\partial t_i} LL\big(\mathbf{x}^{(j+1)}, \mathbf{t}; \breve{\mathbf{y}}\big) = \big(\breve{\mathbf{y}}_i - \mathbf{H}_i(t_i, \mathbf{0})\, \mathbf{x}^{(j+1)}\big)^T \mathbf{\Sigma}_{\mathbf{y}_i}^{-1} \big(\mathbf{D}_i(t_i)\, \mathbf{x}^{(j+1)}\big). \tag{22}$$

As t_is are implicitly involved in the matrices $\mathbf{H}_i(t_i, \mathbf{0})$ and $\mathbf{D}_i(t_i)$, it is not possible to directly compute the roots of the gradient. We use steepest ascend algorithm to update t_i: $t_i^{(j+1)} = t_i^{(j)} + \alpha \frac{\partial}{\partial t_i} LL$. To find the exact step-size, we replace $\mathbf{H}(t_i, \mathbf{0})$ with its first-order Taylor series approximation

$$\mathbf{H}_i(t_i^{(j+1)}, \mathbf{0}) \approx \mathbf{H}_i(t_i^{(j)}, \mathbf{0}) + (\alpha \frac{\partial}{\partial t_i} LL)\, \mathbf{D}_i(t_i^{(j)}). \tag{23}$$

By inserting (23) into LL function of (18) and maximizing it with respect to α,we have

$$\alpha^{(\text{exact})} = \frac{1}{\mathbf{x}^T \mathbf{D}_i(t_i^{(j)})^T\, \mathbf{\Sigma}_{\mathbf{y}_i}^{-1}\, \mathbf{D}_i(t_i^{(j)}) \mathbf{x}} \tag{24}$$

and t_i is updated according to the steepest ascend algorithm as below

$$t_i^{(j+1)} = t_i^{(j)} + \frac{(\breve{\mathbf{y}}_i - \mathbf{H}_i(t_i^{(j)}, \mathbf{0})\mathbf{x})^T\, \mathbf{\Sigma}_{\mathbf{y}_i}^{-1}\, \mathbf{D}_i(t_i^{(j)})\mathbf{x}}{\mathbf{x}^T \mathbf{D}_i(t_i^{(j)})^T\, \mathbf{\Sigma}_{\mathbf{y}_i}^{-1}\, \mathbf{D}_i(t_i^{(j)})\mathbf{x}}, \tag{25}$$

for $1 \le i \le C-1$. As it can be seen from (23), the Taylor expansion is performed around t_i in each iteration of the algorithm. Therefore, we do not need to assume small value for t_i and only the difference between the two estimates $(t_i^{(j+1)} - t_i^{(j)})$ is needed to be small. This enables the algorithm to tolerate large values of time shift errors as demonstrated in Subsection V-A.

C. Initialization and stopping criterion

For the initialization of our iterative method, we assume the ideal case in the beginning. More precisely, we set $t_i^{(0)}$s to zero

$$t_i^{(0)} = 0, \qquad 0 \le i \le C-1, \tag{26}$$

which reflects the timing shift-free setup. In addition, we ignore the jitter noise effect which leads to

$$\mathbf{\Sigma}_{\mathbf{y}_i} = \sigma_w^2\, \mathbf{I}_{N_C}, \qquad 0 \le i \le C-1. \tag{27}$$

According to (15), this corresponds to $\sigma_z = 0$.

The stopping criterion for the algorithm is met when there exists negligible change in the update of the vectors $\mathbf{x}$ and $\mathbf{t}$ ($\|\mathbf{x}^{(j+1)} - \mathbf{x}^{(j)}\|_2^2 < \epsilon_{\mathbf{x}}$ and $\|\mathbf{t}^{(j+1)} - \mathbf{t}^{(j)}\|_2^2 < \epsilon_{\mathbf{t}}$), or the number of iterations exceeds a predefined $\max\_\text{iter}$ number of iteration. Here, $\epsilon_{\mathbf{x}}$ and $\epsilon_{\mathbf{t}}$ are the acceptable tolerance of error for $\mathbf{x}$ and $\mathbf{t}$, respectively. Table I summarizes the proposed approximate iterative ML estimator.

IV. Cramér Rao lower bound

In this section, we present the Cramér-Rao lower bound of the problem (10) for the estimation of the unknown parameter vector

$$\boldsymbol{\alpha} = [t_1, \ldots, t_{C-1}, x_0, \ldots, x_{K-1}]^T. \tag{28}$$

As shown in (17), the likelihood function of $\mathbf{y}$ is decomposed into the multiplication of likelihood functions of $\mathbf{y}_i$s, that are Gaussian random vectors with the mean $\boldsymbol{\eta}_i$ and the covariance matrix $\mathbf{\Sigma}_{\mathbf{y}_i}$. Thus, given the parameters, $\mathbf{y}$ is also a multivariate Gaussian random vector with

$$\mathbf{y} \sim \mathcal{N}(\boldsymbol{\eta}, \mathbf{\Sigma}_{\mathbf{y}}), \tag{29}$$

where $\boldsymbol{\eta} \triangleq [\boldsymbol{\eta}_0^T, \ldots, \boldsymbol{\eta}_{C-1}^T]^T$ is the $N \times 1$ mean vector and $\mathbf{\Sigma}_{\mathbf{y}} \triangleq \text{blkdiag}(\mathbf{\Sigma}_{\mathbf{y}_0}, \ldots, \mathbf{\Sigma}_{\mathbf{y}_{C-1}})$ is the $N \times N$ covariance

TABLE I
ALGORITHM 1: ITERATIVE ML ESTIMATOR FOR THE $\mathbf{x}$ AND $\mathbf{t}$

Input: $\breve{\mathbf{y}}_i$ for $0 \le i \le C-1$, output vectors of each channel.
Output: $\mathbf{x}$, input signal parameters (Nyquist rate samples).
1: Initialize $t_i^{(0)}$ and $\mathbf{\Sigma}_{\mathbf{y}_i}$ for $0 \le i \le C-1$ according to (26) and (27), respectively.
2: $j \leftarrow 0$
3: **repeat**
4: Calculate $\mathbf{x}^{(j+1)}$ using (21).
5: **for** $i = 0$ **to** $C-1$ **do**
6: update $\mathbf{\Sigma}_{\mathbf{y}_i}$ by replacing $\mathbf{x}^{(j+1)}$ and $t_i^{(j)}$ into (15).
7: Calculate $t_i^{(j+1)}$ using (25) ($t_0^{(j+1)} = 0$).
8: update $\mathbf{\Sigma}_{\mathbf{y}_i}$ by replacing $\mathbf{x}^{(j+1)}$ and $t_i^{(j+1)}$ into (15).
9: **end for**
10: **until** $\big(\|\mathbf{x}^{(j+1)} - \mathbf{x}^{(j)}\|_2^2 \ge \epsilon_{\mathbf{x}}$ **or** $\|\mathbf{t}^{(j+1)} - \mathbf{t}^{(j)}\|_2^2 \ge \epsilon_{\mathbf{t}}\big)$ **and** $j \le$ max_iter
11: **return** $\mathbf{x}^{(j+1)}$

matrix ($N = N_C \times C$). Moreover, we define matrices $\mathbf{H}$, $\mathbf{D}$, and $\mathbf{J}$ as

$$\mathbf{H} \triangleq \big[\mathbf{H}_0(t_0,\mathbf{0})^T, \mathbf{H}_1(t_1,\mathbf{0})^T, \ldots, \mathbf{H}_{C-1}(t_{C-1},\mathbf{0})^T\big]^T, \tag{30a}$$

$$\mathbf{D} \triangleq \big[\mathbf{D}_0(t_0)^T, \mathbf{D}_1(t_1)^T, \ldots, \mathbf{D}_{C-1}(t_{C-1})^T\big]^T, \tag{30b}$$

$$\mathbf{J} \triangleq \big[\mathbf{J}_0(t_0)^T, \mathbf{J}_1(t_1)^T, \ldots, \mathbf{J}_{C-1}(t_{C-1})^T\big]^T, \tag{30c}$$

where the $N_C \times K$ matrix $\mathbf{J}_i(t_i)$ is the second derivative of $H_i(t_i,\mathbf{0})$, i.e.,

$$\mathbf{J}_i(t_i)[n,k] = \mathrm{sinc}^{(2)}\left(\tfrac{n}{M} + \tfrac{i}{MC} + t_i - k\right), \tag{31}$$

and $\mathrm{sinc}^{(2)}(x) = \frac{\mathrm{d}^2}{\mathrm{d}x^2}\mathrm{sinc}(x)$.

The Fisher information matrix for the estimation of unknown parameters $\boldsymbol{\alpha}$ from the multivariate Gaussian model of (29) is introduced in [43] as

$$\mathbf{FIM}[i,j] = \frac{1}{2}\,\mathrm{tr}\left\{\frac{\partial \mathbf{\Sigma}_{\mathbf{y}}}{\partial \alpha_i}\mathbf{\Sigma}_{\mathbf{y}}^{-1}\frac{\partial \mathbf{\Sigma}_{\mathbf{y}}}{\partial \alpha_j}\mathbf{\Sigma}_{\mathbf{y}}^{-1}\right\} + \left(\frac{\partial \boldsymbol{\eta}}{\partial \alpha_i}\right)^T \mathbf{\Sigma}_{\mathbf{y}}^{-1}\left(\frac{\partial \boldsymbol{\eta}}{\partial \alpha_j}\right). \tag{32}$$

We use the following Lemma to simplify the calculation of the Fisher information matrix in (32).

Lemma 4.1: Suppose that $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are arbitrary $N \times N$ matrices, $\mathbf{D}_i$ and $\mathbf{E}_j$ are diagonal matrices of size $N \times N$, and $\mathbf{g}_i$ and $\mathbf{h}_j$ are $N \times 1$ vectors, for $0 \le i \le L-1$ and $0 \le j \le M-1$. If we define the $L \times M$ matrix $\mathbf{F}$ as

$$\mathbf{F}[i,j] = \frac{1}{2}\,\mathrm{tr}\{\mathbf{D}_i\,\mathbf{A}\,\mathbf{E}_j\,\mathbf{B}\} + \mathbf{g}_i^T\,\mathbf{C}\,\mathbf{h}_j, \tag{33}$$

the compact form of $\mathbf{F}$ can be obtained as

$$\mathbf{F} = \frac{1}{2}\mathbf{D}^T(\mathbf{A} \odot \mathbf{B}^T)\mathbf{E} + \mathbf{G}^T\mathbf{C}\,\mathbf{H}, \tag{34}$$

where $\mathbf{D}$, $\mathbf{G}$ are $N \times L$ matrices, and $\mathbf{E}$, $\mathbf{H}$ are $N \times M$ matrices with the following definitions

$$\mathrm{diag}(\mathbf{D}[:,i]) = \mathbf{D}_i, \quad \mathbf{G}[:,i] = \mathbf{g}_i,$$
$$\mathrm{diag}(\mathbf{E}[:,j]) = \mathbf{E}_j, \quad \mathbf{H}[:,j] = \mathbf{h}_j,$$

for $0 \le i \le L-1$ and $0 \le j \le M-1$.

Proof: See Appendix. □

The derivative of the vector $\boldsymbol{\eta}$ and the matrix $\mathbf{\Sigma}_{\mathbf{y}}$ with respect to the unknown parameters x_i and t_j for $0 \le i \le K-1$ and $1 \le j \le C-1$ are given by

$$\frac{\partial \boldsymbol{\eta}}{\partial x_i} = \mathbf{H}[:,i], \tag{35}$$

$$\frac{\partial \boldsymbol{\eta_k}}{\partial t_j} = \begin{cases} \mathbf{D}_k(t_k)\mathbf{x} & k = j \\ \mathbf{0} & k \ne j \end{cases},$$

$$\frac{\partial \boldsymbol{\eta}}{\partial t_j} = \mathbf{e}_j \otimes (\mathbf{D}_j(t_j)\mathbf{x}), \tag{36}$$

$$\begin{aligned}\frac{\partial \mathbf{\Sigma}_{\mathbf{y}_k}}{\partial x_i} &= \big\{(\mathbf{D}_k(t_k)\mathbf{x})\mathbf{D}_k(t_k)[:,i]^T \\ &\qquad + \mathbf{D}_k(t_k)[:,i](\mathbf{D}_k(t_k)\mathbf{x})^T\big\} \odot \sigma_z^2\,\mathbf{I}_{N_C} \\ &= \big\{\mathbf{D}_k(t_k)[:,i](\mathbf{D}_k(t_k)\mathbf{x})^T\big\} \odot 2\,\sigma_z^2\,\mathbf{I}_{N_C} \\ &= 2\,\sigma_z^2\,\mathrm{diag}\big\{\mathbf{D}_k(t_k)[:,i] \odot (\mathbf{D}_k(t_k)\mathbf{x})\big\},\end{aligned}$$

$$\frac{\partial \mathbf{\Sigma}_{\mathbf{y}}}{\partial x_i} = 2\,\sigma_z^2\,\mathrm{diag}\big\{\mathbf{D}[:,i] \odot (\mathbf{D}\mathbf{x})\big\}, \tag{37}$$

$$\begin{aligned}\frac{\partial \mathbf{\Sigma}_{\mathbf{y}_j}}{\partial t_j} &= \big\{\mathbf{D}_j(t_j)\mathbf{x}\mathbf{x}^T\mathbf{J}_j^T(t_j) \\ &\qquad + \mathbf{J}_j(t_j)\mathbf{x}\mathbf{x}^T\mathbf{D}_j^T(t_j)\big\} \odot \sigma_z^2\,\mathbf{I}_{N_C} \\ &= \big\{\mathbf{D}_j(t_j)\mathbf{x}\mathbf{x}^T\mathbf{J}_j^T(t_j)\big\} \odot 2\,\sigma_z^2\,\mathbf{I}_{N_C} \\ &= 2\,\sigma_z^2\,\mathrm{diag}\big\{(\mathbf{D}_j(t_j)\mathbf{x}) \odot (\mathbf{J}_j(t_j)\mathbf{x})\big\},\end{aligned}$$

$$\frac{\partial \mathbf{\Sigma}_{\mathbf{y}_k}}{\partial t_j} = \mathbf{0} \qquad k \ne j,$$

$$\frac{\partial \mathbf{\Sigma}_{\mathbf{y}}}{\partial t_j} = 2\,\sigma_z^2\,\mathrm{diag}\big\{\mathbf{e}_j \otimes \big((\mathbf{D}_j(t_j)\mathbf{x}) \odot (\mathbf{J}_j(t_j)\mathbf{x})\big)\big\}, \tag{38}$$

where $\otimes$ represents the Kronecker product and $\mathbf{e}_j$ is a vector with one in jth position and zero in the other positions.

We utilize Lemma *4.1* together with equations (35) and (37) to calculate the Fisher information matrix related to vector $\mathbf{x}$

$$\begin{aligned}\mathbf{FIM}_{\mathbf{xx}} &= \frac{1}{2}\,\mathbf{F}_{\mathbf{x}}^T(\mathbf{\Sigma}_{\mathbf{y}}^{-1} \odot \mathbf{\Sigma}_{\mathbf{y}}^{-T})\mathbf{F}_{\mathbf{x}} + \mathbf{H}^T\mathbf{\Sigma}_{\mathbf{y}}^{-1}\mathbf{H} \\ &= \frac{1}{2}\,\mathbf{F}_{\mathbf{x}}^T\,\mathbf{\Sigma}_{\mathbf{y}}^{-2}\,\mathbf{F}_{\mathbf{x}} + \mathbf{H}^T\mathbf{\Sigma}_{\mathbf{y}}^{-1}\mathbf{H},\end{aligned} \tag{39}$$

where $\mathbf{F}_{\mathbf{x}}[:,i] = 2\,\sigma_z^2\big(\mathbf{D}[:,i] \odot (\mathbf{D}\mathbf{x})\big)$ for $0 \le i \le K-1$. The latter expression in (39) comes from the diagonal property of matrix $\mathbf{\Sigma}_{\mathbf{y}}$. Similarly, by using (36) and (38), we have

$$\mathbf{FIM}_{\mathbf{tt}} = \frac{1}{2}\,\mathbf{F}_{\mathbf{t}}^T\,\mathbf{\Sigma}_{\mathbf{y}}^{-2}\,\mathbf{F}_{\mathbf{t}} + \mathbf{M}^T\mathbf{\Sigma}_{\mathbf{y}}^{-1}\mathbf{M}, \tag{40}$$

$$\mathbf{FIM}_{\mathbf{xt}} = \frac{1}{2}\,\mathbf{F}_{\mathbf{x}}^T\,\mathbf{\Sigma}_{\mathbf{y}}^{-2}\,\mathbf{F}_{\mathbf{t}} + \mathbf{H}^T\mathbf{\Sigma}_{\mathbf{y}}^{-1}\mathbf{M}, \tag{41}$$

$$\mathbf{FIM}_{\mathbf{tx}} = \mathbf{FIM}_{\mathbf{xt}}^T. \tag{42}$$

where

$$\mathbf{M}[:,j] = \mathbf{e}_j \otimes (\mathbf{D}_j(t_j,\mathbf{0})\mathbf{x}), \tag{43}$$

$$\mathbf{F}_{\mathbf{t}}[:,j] = 2\,\sigma_z^2\big\{\mathbf{e}_j \otimes \big((\mathbf{D}_j(t_j,\mathbf{0})\mathbf{x}) \odot (\mathbf{J}_j(t_j,\mathbf{0})\mathbf{x})\big)\big\} \tag{44}$$

for $1 \le j \le C-1$.

The $\mathbf{FIM}$ matrix can be obtained by putting the submatrices of (39) to (42) together, i.e.,

$$\mathbf{FIM} = \begin{bmatrix} \mathbf{FIM}_{\mathbf{xx}} & \mathbf{FIM}_{\mathbf{xt}} \\ \mathbf{FIM}_{\mathbf{tx}} & \mathbf{FIM}_{\mathbf{tt}} \end{bmatrix}. \tag{45}$$

From the Cramér Rao Theorem, the following inequality is achieved as the lower bound for the MSE of vector $\mathbf{x}$:

$$\mathbb{E}[(\mathbf{x}-\hat{\mathbf{x}})^T(\mathbf{x}-\hat{\mathbf{x}})] \succcurlyeq \mathrm{tr}\{\mathbf{C_{xx}}\}, \tag{46}$$

where

$$\mathbf{FIM}^{-1} = \begin{bmatrix} \mathbf{C_{xx}} & \mathbf{C_{xt}} \\ \mathbf{C_{tx}} & \mathbf{C_{tt}} \end{bmatrix}, \tag{47}$$

and $\mathbf{C_{xx}}$, $\mathbf{C_{xt}}$, $\mathbf{C_{tx}}$, and $\mathbf{C_{tt}}$ are the sub-matrices of $\mathbf{FIM}^{-1}$ with the same dimensions as $\mathbf{FIM_{xx}}$, $\mathbf{FIM_{xt}}$, $\mathbf{FIM_{tx}}$, and $\mathbf{FIM_{tt}}$, respectively.

V. Numerical Simulations

In this section, the behavior of our iterative ML estimator is presented. We compare the performance of our algorithm with its Cramér-Rao lower bound and the linear estimator with no compensation of the timing mismatch and jitter noise. Moreover, we study the bias of the proposed algorithm and its sensitivity to erroneous values of jitter and additive noises' variances.

A. Performance of the Iterative ML algorithm

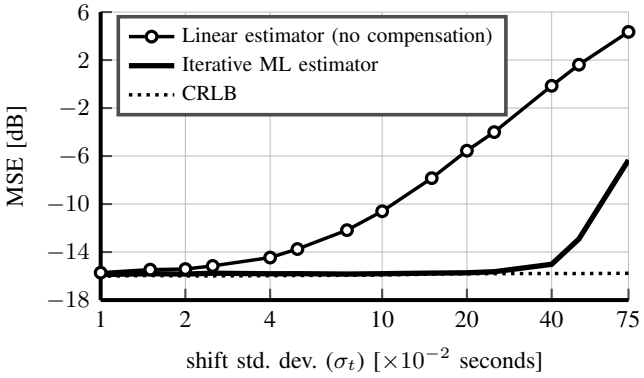
We use the mean square error (MSE) as the measurement of performance for the signal parameters estimation in our experiments. For the simulations, 10000 independent runs are executed, and each point in the figures represents the average of square error for these runs. Using MATLAB, we generate random jitter and additive noises with the zero mean Gaussian distributions of variances σ_z^2 and σ_w^2, respectively. The deterministic variables of the signal parameters, x_k $0 \le k \le K-1$, and the shift mismatches, t_i $1 \le i \le C-1$, also have zero mean Gaussian distributions with variances of σ_x^2 and σ_t^2, respectively. In addition, we choose $K = 10$, $\sigma_w = 0.05$, and $\sigma_x = 1$. The stopping criteria for $\epsilon_{\mathbf{x}}$ and $\epsilon_{\mathbf{t}}$ are set to 10^{-8} and $\max\_\mathrm{iter} = 1000$.

In the absence of time shift errors and jitter noise we have $\mathbf{z}_i = \mathbf{0}$ and $t_i = 0$, which yields $\mathbf{H}_i(t_i, \mathbf{z}_i) = \mathbf{H}(0, \mathbf{0})$ and $\mathbf{\Sigma}_{\mathbf{y}_i} = \sigma_w^2 \mathbf{I}_{N_C}$ for $0 \le i \le C-1$. Hence, the output vectors $\mathbf{y}_i$ simply can be expressed as a linear relationship with additive i.i.d. Gaussian noise, i.e. $\mathbf{y}_i = \mathbf{H}_i(0,\mathbf{0})\,\mathbf{x} + \mathbf{w}_i$. The linear estimator for $\mathbf{x}$, in this case, is obtain as

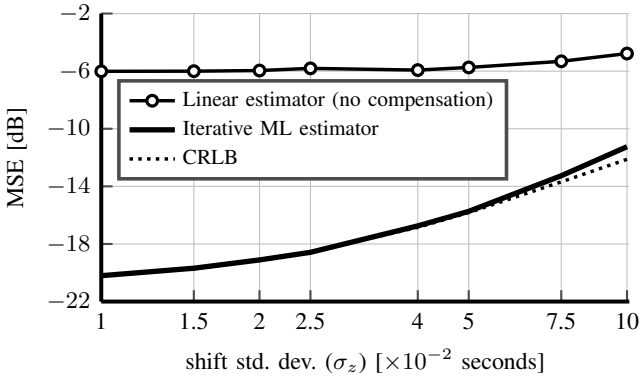
$$\hat{\mathbf{x}} = \Big[\sum_{i=0}^{C-1} \mathbf{H}_i(0,\mathbf{0})^T \mathbf{H}_i(0,\mathbf{0})\Big]^{-1} \sum_{i=0}^{C-1} \mathbf{H}_i(0,\mathbf{0})^T \breve{\mathbf{y}}_i. \tag{48}$$

Here for Gaussian assumption, the best linear unbiased estimator (BLUE) of (48) is also the minimum variance unbiased estimator (MVUE). This means that the linear uncompensated estimator of (48) is the best estimator among all other estimators in the MSE sense when we have no time shift errors and jitter noise.

In Fig. 2, the iterative ML algorithm is compared against the Cramér-Rao lower bound and the linear uncompensated estimator for the various values of σ_z and σ_t. The number of channels is $C = 4$ and the oversampling ratio is $M = 1$. As we can see from the curves in Fig. 2, the iterative ML algorithm almost follows its CRLB for both of the scenarios.



(a) different values of σ_t for $\sigma_z = 0.05$



(b) different values of σ_z for $\sigma_t = 0.2$

Fig. 2. MSE performance of iterative ML estimator for different values of σ_t and σ_z, with $\sigma_w = 0.05$, $M = 1$, and $C = 4$.

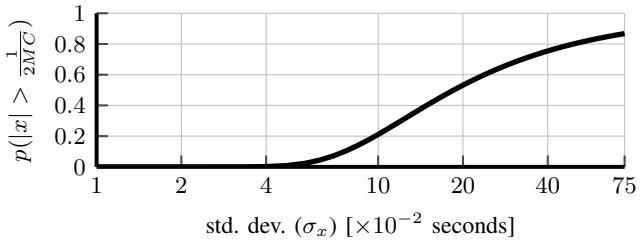
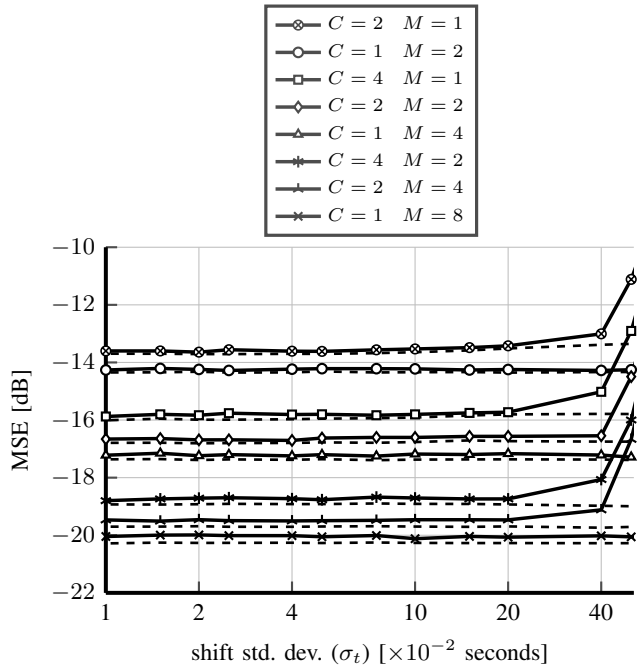


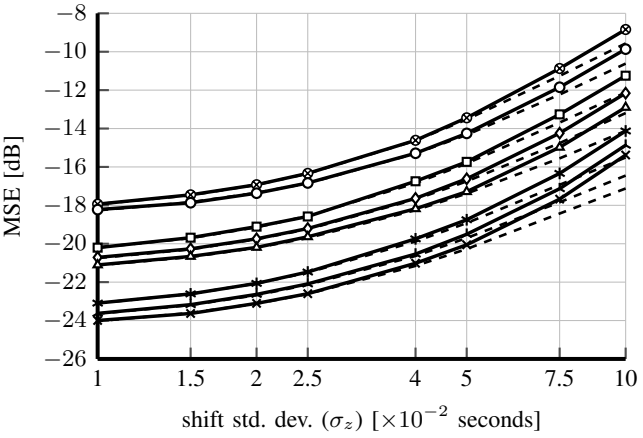
Fig. 3. The probability that the sampling time goes beyond its nominal region, i.e. $p\{|x| > \frac{1}{2MC}\}$.

The MSE of our algorithm starts to separate from CRLB for $\sigma_z > 0.05$. The probability that the sampling time goes beyond its nominal region, i.e. $p\{|x| > \frac{1}{2MC}\}$, is depicted in Fig. 3 for the zero mean Gaussian variable and different standard deviations. This figure shows that we begin to have violation on the ordering of samples which is rare to be happened in practice. The iterative ML algorithm operates near CRLB even for larger values of σ_t rather than σ_z. This resistance to high values of shift mismatches is more interesting for us since it is reasonable to have bigger shift mismatches in comparison with jitter deviations.

The next performance experiment describes the multichannel effect for different cases of M and C. It is important in each multichannel signal acquisition framework to compare its performance with a higher rate sampler in order to measure the performance loss due to the channel mismatches. Fig. 4 depicts



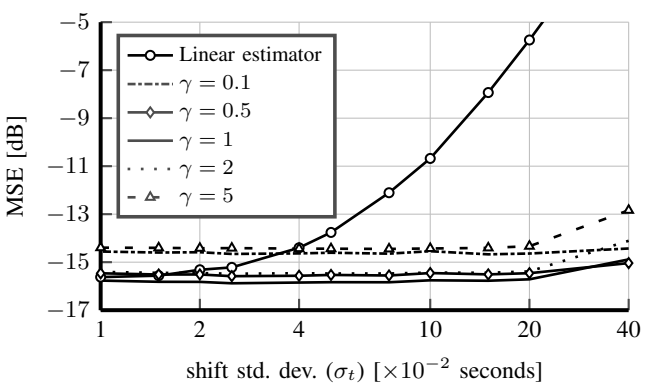
(a) different values of σ_t for $\sigma_z = 0.05$



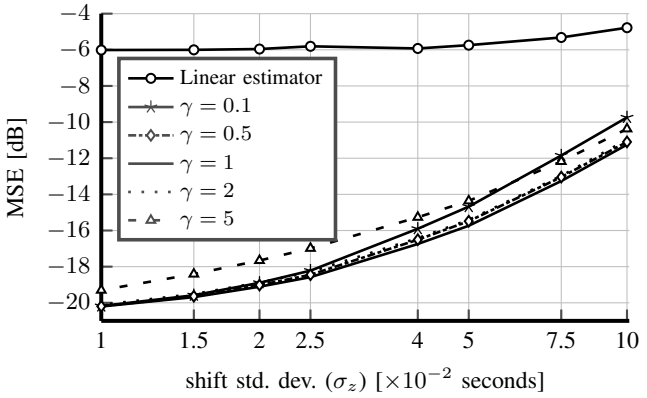
(b) different values of σ_z for $\sigma_t = 0.2$

Fig. 4. Comparison of iterative ML estimator method between the samplers with $MC = 2$, $MC = 4$, and $MC = 8$, for $\sigma_w = 0.05$ and different values of σ_t and σ_z (The legend for Fig. 4a and Fig. 4b is shown in the upper box).

three overall sampling rate group, $MC = 2$, $MC = 4$, and $MC = 8$, with their corresponding CRLBs shown with dashed lines under each sampler's curve. It can be seen from Fig. 4a and Fig. 4b that the samplers with the same product of $M \times C$ have near MSE performances and the samplers with multiple channels perform almost closely to the sampler of the same product with one channel yet higher sampling rate. Whereas according to Fig. 2, using multichannel structure without compensation cannot demonstrate its superiority even over the samplers with lower rates. For instance, linear uncompensated sampler of $M = 1$, $C = 4$ has -10 dB MSE in $\sigma_z = 0.05$ and $\sigma_t = 0.1$, where it is -14.5 dB in the sampler with $M = 2$, $C = 1$.



(a) different values of σ_t for $\sigma_z = 0.05$



(b) different values of σ_z for $\sigma_t = 0.2$

Fig. 5. Sensitivity evaluation of the iterative ML method to the estimation errors of the variances ratio of $(\frac{\sigma_z}{\sigma_w})^{(est.)} = \gamma(\frac{\sigma_z}{\sigma_w})$ for cases $\gamma = 0.1$, $\gamma = 0.5$, $\gamma = 1$, $\gamma = 2$, and $\gamma = 5$, with $\sigma_w = 0.05$, $M = 1$, and $C = 4$.

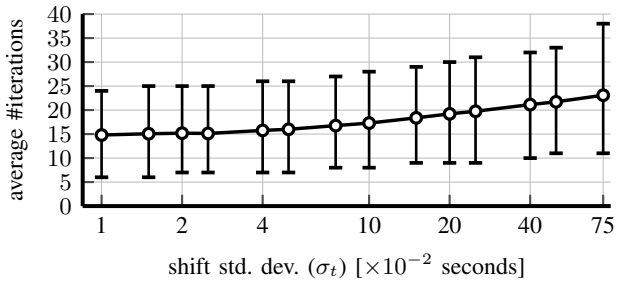
B. Sensitivity to errors in σ_z^2 and σ_w^2

In this subsection, the sensitivity of the proposed algorithm to the errors in σ_z^2 and σ_w^2 is analyzed. We suppose that the σ_z^2 and σ_w^2 are estimated beforehand. Therefore, we measure MSE performance of the method in the cases that the input values of the algorithm for these variances differ from the original ones. As it can be found from (21) and (25), multiplication of $\mathbf{\Sigma}_{\mathbf{y}_i}$s by a constant does not affect formulas (21) and (25). Multiplying $\mathbf{\Sigma}_{\mathbf{y}_i}$s with $\frac{1}{\sigma_w^2}$ in (15) helps that the algorithm only depends on the proportion of $\frac{\sigma_z}{\sigma_w}$. Fig. 5 plots the MSE for various values of γ where $(\frac{\sigma_z}{\sigma_w})^{(est.)} = \gamma\frac{\sigma_z}{\sigma_w}$ is given to the algorithm. The difference between the original and the given portions in the variances causes the performance loss in the algorithm. For the cases of $\gamma = 0.1$ and $\gamma = 5$, the algorithm's MSE becomes even higher than the linear estimator for small amounts of σ_t ($\sigma_t < 0.04$). However, slight increase in σ_t results in good distance between the curves with $\gamma \neq 1$ and the uncompensated one.

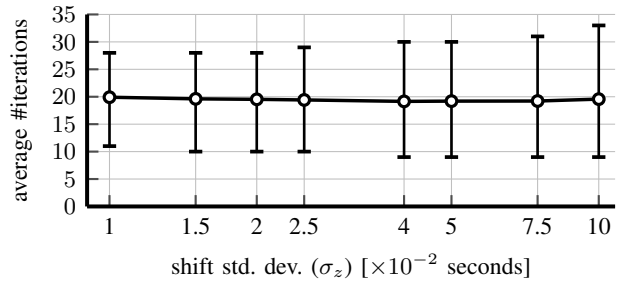
VI. Discussion

A. Computational complexity

The computational complexity of the proposed method can be written as the number of iterations the algorithm needs to converge multiplied with the number of operations required in



(a) different values of σ_t for $\sigma_z = 0.05$



(b) different values of σ_z for $\sigma_t = 0.2$

Fig. 6. Average number of iterations that iterative ML algorithm needs to converge for different values of σ_t and σ_z, with $\sigma_w = 0.05$, $M = 1$, and $C = 4$. The error bar indicates the interval that includes 95% of data points around the average point.

TABLE II
COMPUTATIONAL COMPLEXITY FOR ONE ITERATION OF OUR METHOD FOR TWO CASES OF CORRELATED AND UNCORRELATED JITTER NOISE. THE COST OF EACH PART OF THE ALGORITHM IS PRESENTED SEPARATELY.

	correlated jitter noise	uncorrelated jitter noise
calculate $\mathbf{\Sigma}_{\mathbf{y}_i}$s	$\mathcal{O}(C\,N_C^2 + K\,N_C)$	$\mathcal{O}(C\,K\,N_C)$
update $\mathbf{x}$	$\mathcal{O}(C\,N_C^3 + C\,K\,N_C^2 + C\,K^2\,N_C + K^3)$	$\mathcal{O}(K^3 + C\,K^2\,N_C)$
update t_is	$\mathcal{O}(C\,N_C^3)$	$\mathcal{O}(C\,K\,N_C)$

each iteration. Fig. 6 displays the average number of iterations needed for convergence of the algorithm in 10000 independent trials. The error bars show the intervals in which 95% of data points exist. On average, we observe that increasing σ_t causes slower converge of the algorithm. However, the average number of iterations is less than 25 for a broad range of $\sigma_t < 0.75$. This number is almost constant for different values of σ_z.

Next, we analyze the complexity of each iteration. In this analysis, N_C and K are assumed to be large variables with the same growing rate. In each iteration of the algorithm, we calculate $\mathbf{\Sigma}_{\mathbf{y}_i}$s twice and updating $\mathbf{x}$ and t_is once. The complexity of these different parts are presented in Table II for both correlated an uncorrelated jitter noise. Since in uncorrelated case the covariance matrices $\mathbf{\Sigma}_{\mathbf{y}_i}$s become diagonal, the required numbers of operations are greatly reduce compared with the correlated case. According to Table II, updating $\mathbf{x}$ is the most costly part among others.

B. Sub-Nyquist sampling scenario

For our method, we need to have $M\,C \geq 1$ to avoid underdetermined system and choose M such that the number of samples collected by each channel belongs to positive integer, i.e. $N_C = M\,K \in \mathbb{N}$. Moreover, in order to have full-rank matrices of $\mathbf{H}_i(t_i, \mathbf{0})^T \mathbf{\Sigma}_{\mathbf{y}_i}^{-1} \mathbf{H}_i(t_i, \mathbf{0})$, M cannot take values less than 1. Otherwise, one may find particular set of $\{t_i\}_{i=1}^{C-1}$ that causes the following matrix

$$\sum_{i=0}^{C-1} \mathbf{H}_i(t_i, \mathbf{0})^T \mathbf{\Sigma}_{\mathbf{y}_i}^{-1} \mathbf{H}_i(t_i, \mathbf{0}) \tag{49}$$

in equation (21) not to be invertible. For example, a prevalent case observed in simulations is when t_is tend to $\frac{-i}{M\,C}$ in the iterative algorithm. In this case, matrices $\mathbf{H}_i(t_i, \mathbf{0})$ and $\mathbf{\Sigma}_{\mathbf{y}_i}$ approaches $\mathbf{H}(0, \mathbf{0})$ and $\mathbf{\Sigma}_{\mathbf{y}_0}$ for $1 \leq i \leq C-1$, respectively, and the summation of (49) goes towards the singular matrix of $C\,\mathbf{H}_0(0, \mathbf{0})^T \mathbf{\Sigma}_{\mathbf{y}_0}^{-1} \mathbf{H}(0, \mathbf{0})$. Therefore, the method fails to estimate the signal parameters correctly. Finding stable algorithms that is able to operate in sub-Nyquist sampling case ($M < 1$) is a topic of future works.

C. Practical remarks

From a practical point of view, the proposed method can offer attractive features. It may not possible or cost-efficient in some applications to change the analog hardware such as ADCs, but instead it is feasible to utilize different digital processing algorithms by updating a microcontroller or adding a new processing unit. In some approaches like mixed-signal techniques, it is required to design a new ADC architecture with adjustable analog circuits to eliminate the effects of time shift error. However, our method is able to work with different types of conventional ADCs. Another advantage of the proposed method is its capability of estimating the time shift errors without the need for an additional reference ADC channel, as opposed to many correlation-based algorithm. Moreover, since the proposed algorithm is blind the acquisition process has not to be periodically interrupted for calibration. One of the main practical advantage of this method is for applications that off-line post-processing of the collected data is possible. In these applications, we can potentially reduce power consumption of on-chip acquisition circuits by utilizing low-power TI-ADC with greater jitter noise and perform the compensation task with iterative ML algorithm. In recent years, the performance and efficiency of integrated circuits have significantly improved. Therefore, it could be possible to implement the algorithm based on application specific integrated circuit (ASIC) which allows block-processing of ADC outputs in real-time.

VII. CONCLUSION

The TI-ADC is a low-cost analog to digital structure that can achieve high power efficiency in sampling. However, its performance is greatly affected by timing mismatch. Therefore, it is essential to estimate and compensate the time shift errors. Moreover, canceling the jitter noise through digital processing can bring additional reduction in total power by eliminating

the utilization of high power, accurate clock sources. In this paper, we proposed a blind technique to mitigate the jitter noise and the timing mismatch. The method employs an iterative ML estimator to obtain the time shift errors, and eventually achieve the uniform samples of the signal. Although we use some simplifying approximations for deriving our algorithm, the simulation results reveal that the performance of the method tightly follows its CRLB for σ_t and σ_z lower than 1.6 and 0.2 of the sampling time (0.25 seconds for the case of $C = 4$ and $M = 1$), respectively. Fortunately, such large values for variances of the shift errors and jitter noise are uncommon in most of the practical cases. Furthermore, we can see significant improvement over the linear estimation case where no compensation is applied. Our method also has lower estimation bias compared with linear estimator. For performing the iterative ML estimator, the ratio of variances $(\frac{\sigma_z}{\sigma_w})$ is required. However, this method is capable of tolerating relatively large amount of errors in the estimation of this ratio.

APPENDIX
PROOF OF LEMMA *4.1*

For the first term in (33), we can write

$$\begin{aligned}
\mathrm{tr}\{\mathbf{D}_i\mathbf{A}\mathbf{E}_j\mathbf{B}\} &= \sum_{n=0}^{N-1}(\mathbf{D}_i\mathbf{A}\mathbf{E}_j\mathbf{B})[n,n] \\
&= \sum_{n=0}^{N-1}\sum_{k=0}^{N-1}(\mathbf{D}_i\mathbf{A})[n,k](\mathbf{E}_j\mathbf{B})[k,n] \\
&= \sum_{n=0}^{N-1}\sum_{k=0}^{N-1}\mathbf{D}_i[n,n]\mathbf{A}[n,k]\mathbf{E}_j[k,k]\mathbf{B}[k,n] \\
&= \sum_{n=0}^{N-1}\sum_{k=0}^{N-1}\mathbf{D}[n,i]\mathbf{A}[n,k]\mathbf{E}[k,j]\mathbf{B}[k,n] \\
&= \sum_{n=0}^{N-1}\sum_{k=0}^{N-1}\mathbf{D}^T[i,n](\mathbf{A}\odot\mathbf{B}^T)[n,k]\mathbf{E}[k,j] \\
&= \big(\mathbf{D}^T(\mathbf{A}\odot\mathbf{B}^T)\mathbf{E}\big)[i,j]. \qquad (50)
\end{aligned}$$

The second term is also compacted as

$$\begin{aligned}
\mathbf{g}_i^T\mathbf{C}\mathbf{h}_j &= \sum_{n=0}^{N-1}\sum_{k=0}^{N-1}\mathbf{g}_i[n,1]\mathbf{C}[n,k]\mathbf{h}_j[k,1] \\
&= \sum_{n=0}^{N-1}\sum_{k=0}^{N-1}\mathbf{G}[n,i]\mathbf{C}[n,k]\mathbf{H}[k,j] \\
&= \big(\mathbf{G}^T\mathbf{C}\mathbf{H}\big)[i,j]. \qquad (51)
\end{aligned}$$

Finally, we can achieve the equation (34) by replacing (50) and (51) into (33).

REFERENCES

[1] A. A. Abidi, "The path to the software-defined radio receiver," *IEEE J. Solid-State Circuits*, vol. 42, pp. 954–966, May 2007.

[2] A. K. Salkintzis, H. Nie, and P. T. Mathiopoulos, "ADC and DSP challenges in the development of software radio base stations," *IEEE Pers. Commun.*, vol. 6, pp. 47–55, Aug. 1999.

[3] R. Blazquez, P. P. Newaskar, F. S. Lee, and A. P. Chandrakasan, "A baseband processor for impulse ultra-wideband communications," *IEEE J. Solid-State Circuits*, vol. 40, pp. 1821–1828, Sep. 2005.

[4] J. Singh, S. Ponnuru, and U. Madhow, "Multi-gigabit communication: the ADC bottleneck," in *Proc. IEEE International Conference on Ultra-Wideband*, Sep. 2009, pp. 22–27.

[5] E. Axell, G. Leus, E. G. Larsson, and H. V. Poor, "Spectrum sensing for cognitive radio : State-of-the-art and recent advances," *IEEE Signal Process. Mag.*, vol. 29, pp. 101–116, May 2012.

[6] H. Sun, A. Nallanathan, C. X. Wang, and Y. Chen, "Wideband spectrum sensing for cognitive radio networks: a survey," *IEEE Wireless Commun.*, vol. 20, pp. 74–81, Apr. 2013.

[7] B. Murmann, "A/D converter trends: Power dissipation, scaling and digitally assisted architectures," in *Proc. IEEE Custom Integrated Circuits Conference*, Sep. 2008, pp. 105–112.

[8] W. C. Black and D. A. Hodges, "Time interleaved converter arrays," *IEEE J. Solid-State Circuits*, vol. 15, pp. 1022–1029, Dec. 1980.

[9] B. Razavi, "Design considerations for interleaved ADCs," *IEEE J. Solid-State Circuits*, vol. 48, pp. 1806–1817, Aug. 2013.

[10] Y.-C. Jenq, "Digital spectra of nonuniformly sampled signals: fundamentals and high-speed waveform digitizers," *IEEE Trans. Instrum. Meas.*, vol. 37, pp. 245–251, Jun. 1988.

[11] K. C. Dyer, D. Fu, S. H. Lewis, and P. J. Hurst, "An analog background calibration technique for time-interleaved analog-to-digital converters," *IEEE J. Solid-State Circuits*, vol. 33, pp. 1912–1919, Dec. 1998.

[12] H. Masuda, S. Ohkawa, A. Kurokawa, and M. Aoki, "Challenge: variability characterization and modeling for 65- to 90-nm processes," in *Proc. IEEE Custom Integrated Circuits Conference*, Sep. 2005, pp. 593–599.

[13] A. Petraglia and S. K. Mitra, "Analysis of mismatch effects among a/d converters in a time-interleaved waveform digitizer," *IEEE Trans. Instrum. Meas.*, vol. 40, pp. 831–835, Oct. 1991.

[14] N. Kurosawa, H. Kobayashi, K. Maruyama, H. Sugawara, and K. Kobayashi, "Explicit analysis of channel mismatch effects in time-interleaved ADC systems," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 48, pp. 261–271, Mar. 2001.

[15] C. Vogel, "The impact of combined channel mismatch effects in time-interleaved ADCs," *IEEE Trans. Instrum. Meas.*, vol. 54, pp. 415–427, Feb. 2005.

[16] J. M. D. Pereira, P. M. B. S. Girao, and A. M. C. Serra, "An FFT-based method to evaluate and compensate gain and offset errors of interleaved ADC systems," *IEEE Trans. Instrum. Meas.*, vol. 53, pp. 423–430, Apr. 2004.

[17] V. Ferragina, A. Fornasari, U. Gatti, P. Malcovati, and F. Maloberti, "Gain and offset mismatch calibration in time-interleaved multipath A/D sigma-delta modulators," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 51, pp. 2365–2373, Dec. 2004.

[18] A. Biallais and F. Ghanem, "Systems and methods for estimation of offset and gain errors in a time-interleaved analog-to-digital converter," U.S. Patent 9 154 147, Oct. 6, 2015.

[19] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas Commun.*, vol. 17, pp. 539–550, Apr. 1999.

[20] B. Brannon, "Sampled systems and the effects of clock phase noise and jitter," Analog Devices, Inc., Norwood, MA, Application Note AN-756, 2004.

[21] T. H. Lee and A. Hajimiri, "Oscillator phase noise: a tutorial," *IEEE J. Solid-State Circuits*, vol. 35, pp. 326–336, Mar. 2000.

[22] K. Uyttenhove and M. S. J. Steyaert, "Speed-power-accuracy tradeoff in high-speed CMOS ADCs," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 49, pp. 280–287, Apr. 2002.

[23] D. S. Weller, "Mitigating timing noise in ADCs through digital post-processing," Master's thesis, Massachusetts Institute of Technology, Cambridge, 2008.

[24] J. Elbornsson, F. Gustafsson, and J. E. Eklund, "Blind equalization of time errors in a time-interleaved adc system," *IEEE Trans. Signal Process.*, vol. 53, pp. 1413–1424, Apr. 2005.

[25] P. Vandewalle, L. Sbaiz, J. Vandewalle, and M. Vetterli, "Super-resolution from unregistered and totally aliased signals using subspace methods," *IEEE Trans. Signal Process.*, vol. 55, pp. 3687–3703, Jul. 2007.

[26] V. Divi and G. Wornell, "Bandlimited signal reconstruction from noisy periodic nonuniform samples in time-interleaved ADCS," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Las Vegas, NV, Mar. 2008, pp. 3721–3724.

[27] J. Goodman, B. Miller, M. Herman, G. Raz, and J. Jackson, "Polyphase nonlinear equalization of time-interleaved analog-to-digital converters," *IEEE J. Sel. Topics Signal Process.*, vol. 3, pp. 362–373, Jun. 2009.

[28] H. Johansson, "A polynomial-based time-varying filter structure for the compensation of frequency-response mismatch errors in time-interleaved

ADCs," *IEEE J. Sel. Topics Signal Process.*, vol. 3, pp. 384–396, Jun. 2009.

[29] V. Divi and G. W. Wornell, "Blind calibration of timing skew in time-interleaved analog-to-digital converters," *IEEE J. Sel. Topics Signal Process.*, vol. 3, pp. 509–522, Jun. 2009.

[30] S. Singh, L. Anttila, M. Epp, W. Schlecker, and M. Valkama, "Frequency response mismatches in 4-channel time-interleaved ADCs: analysis, blind identification, and correction," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 62, pp. 2268–2279, Sep. 2015.

[31] Y. X. Zou, S. L. Zhang, Y. C. Lim, and X. Chen, "Timing mismatch compensation in time-interleaved adcs based on multichannel lagrange polynomial interpolation," *IEEE Trans. Instrum. Meas.*, vol. 60, pp. 1123–1131, Apr. 2011.

[32] C. A. Schmidt, J. E. Cousseau, J. L. Figueroa, B. T. Reyes, and M. R. Hueda, "Efficient estimation and correction of mismatch errors in time-interleaved ADCs," *IEEE Trans. Instrum. Meas.*, vol. 65, pp. 243–254, Feb. 2016.

[33] A. Balakrishnan, "On the problem of time jitter in sampling," *IRE Trans. Inf. Theory*, vol. 8, pp. 226–236, Apr. 1962.

[34] J. Verspecht, "Compensation of timing jitter-induced distortion of sampled waveforms," *IEEE Trans. Instrum. Meas.*, vol. 43, pp. 726–732, Oct. 1994.

[35] G. Vandersteen and R. Pintelon, "Maximum likelihood estimator for jitter noise models," *IEEE Trans. Instrum. Meas.*, vol. 49, pp. 1282–1284, Dec. 2000.

[36] A. Nordio, C.-F. Chiasserini, and E. Viterbo, "Signal reconstruction errors in jittered sampling," *IEEE Trans. Signal Process.*, vol. 57, pp. 4711–4718, Dec 2009.

[37] A. Kumar, "On bandlimited signal reconstruction from the distribution of unknown sampling locations," *IEEE Trans. Signal Process.*, vol. 63, pp. 1259–1267, Mar. 2015.

[38] D. S. Weller and V. K. Goyal, "Bayesian post-processing methods for jitter mitigation in sampling," *IEEE Trans. Signal Process.*, vol. 59, pp. 2112–2123, May 2011.

[39] ——, "On the estimation of nonrandom signal coefficients from jittered samples," *IEEE Trans. Signal Process.*, vol. 59, pp. 587–597, Feb. 2011.

[40] S. Kumar, V. K. Goyal, and S. E. Sarma, "Efficient parametric signal estimation from samples with location errors," *IEEE Trans. Signal Process.*, vol. 61, pp. 5285–5297, Nov. 2013.

[41] H. Araghi, M. A. Akhaee, and A. Amini, "Timing mismatch compensation in TI-ADCs using Bayesian approach," in *23rd European Signal Processing Conference (EUSIPCO)*, Nice, France, Aug. 2015, pp. 1391–1395.

[42] ——, "Approximate ML estimator for compensation of timing mismatch and jitter noise in TI-ADCs," in *24th European Signal Processing Conference (EUSIPCO)*, Budapest, Hungary, Aug. 2016, pp. 2360–2364.

[43] P. Stoica and A. Nehorai, "Music, maximum likelihood, and Cramer-Rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 720–741, May 1989.