# Low-Sampling-Rate Ultra-Wideband Channel Estimation Using Equivalent-Time Sampling

Tarig Ballal and Tareq Y. Al-Naffouri, *Member, IEEE*

***Abstract*—In this paper, a low-sampling-rate scheme for ultra-wideband channel estimation is proposed. The scheme exploits multiple observations generated by transmitting multiple pulses. In the proposed scheme, $P$ pulses are transmitted to produce channel impulse response estimates at a desired sampling rate, while the ADC samples at a rate that is $P$ times slower. To avoid loss of fidelity, the number of sampling periods (based on the desired rate) in the inter-pulse interval is restricted to be co-prime with $P$. This condition is affected when clock drift is present and the transmitted pulse locations change. To handle this case, and to achieve an overall good channel estimation performance, without using prior information, we derive an improved estimator based on the *bounded data uncertainty* (BDU) model. It is shown that this estimator is related to the Bayesian *linear minimum mean squared error* (LMMSE) estimator. Channel estimation performance of the proposed sub-sampling scheme combined with the new estimator is assessed in simulation. The results show that high reduction in sampling rate can be achieved. The proposed estimator outperforms the *least squares* estimator in almost all cases, while in the high SNR regime it also outperforms the LMMSE estimator. In addition to channel estimation, a synchronization method is also proposed that utilizes the same pulse sequence used for channel estimation.**

***Index Terms*—Ultra-wideband, UWB, channel estimation, sub-sampling, ADC, linear minimum mean squared error, LMMSE, bounded data uncertainty, BDU.**

## I. Introduction

CHANNEL estimation is an important process for ultra-wideband (UWB) short-range communication and high precision location and navigation systems [1], [2]. The main challenge in the problem stems from the large bandwidth of the transmitted pulses, which results in an extravagantly high *Nyquist* sampling rate (equivalent to twice the bandwidth), thus leading to formidable *analogue-to-digital converter* (ADC) requirements [1]–[3]. Due to the same large bandwidth, a large number of multipath echoes are resolvable [1]. As a result, an even higher (than Nyquist) rate might be required for resolving these multipath components. For example, in [4], a sampling rate in the range of 17.9–35.7 GHz was suggested in the context of UWB channel estimation. Sampling at such a rate is practically limited by the cost and complexity of the required hardware [1].

Techniques like *compressed sensing* are commonly applied to reduce the ADC sampling rate required for signal reconstruction and/or parameterhe estimation [5]. These techniques capitalize on the sparsity of the data of interest. Due to the high density of UWB multipath in most practical situations, these techniques are not of much use in the UWB case. For example, according to the IEEE 802.15.3a standard [6], the $\mathrm{CM}_1$ model (based on line-of-sight channel measurements at 0–4 m) has a cluster arrival rate of 0.0233 $\mathrm{ns}^{-1}$ and ray arrival rate of 2.5 $\mathrm{ns}^{-1}$ within each cluster. Considering one of these clusters, we have a duration of approximately 43 ns in which we have, approximately, 107 distinct multipath components. To resolve all these components, we need to sample at $f_s \geq 107/(43\times 10^{-9}) \approx 2.5$ GHz, in which case the sparsity rate is 1 (out of 1). For the channel to be considered sufficiently sparse (e.g., has a sparsity rate of 0.1 or less) the channel has to be sampled at a sampling rate $\mathrm{f_s} \geq (107/0.1)/(43\times 10^{-9}) \approx 25$ GHz. This raises a conflict between the sampling requirement for UWB signals in order to make them sparse on the one hand, and the general interest in reducing the cost of the sampling hardware on the other hand. It can be said, therefore, that the use of compressed sensing will not be efficient with UWB channels, at least as far as indoor environments are concerned.

Persistent efforts have been made to reduce the sampling resources required for UWB channel estimation. For example, in [2], a dictionary of parameterized waveforms (atoms) is designed such that a sparse representation of the received signal can be assumed. A compressed-sensing technique is applied to reduce the sampling rate to 1/3 of the Nyquist rate. Although this is a significant reduction in sampling rate, the required sampling rate is still expensive for commercial applications. Improving this technique is practically limited by the sparsity rates of the UWB channels, as has been alluded to in the above discussion.

In [7], the channel estimation problem is translated into a harmonic retrieval problem to achieve up to 1/8 sampling rate reduction. However, as is explained in [1], this method is blind to circular shifts, and therefore, it cannot estimate timing offsets. Unlike [7], in [8], a separate timing estimator is combined with the frequency-domain channel estimation method to offer a more robust alternative. Unfortunately, the high sampling requirements of the latter approach make the approach inefficient. In [9], a bank of parallel analogue matched filters is used to

Manuscript received November 19, 2013; revised March 30, 2014; accepted June 07, 2014. Date of publication July 17, 2014; date of current version August 20, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Adel Belouchrani.

T. Ballal is with the Electrical Engineering Department, King Abdullah University of Science and Technology (KAUST), Thuwal 23955, Saudi Arabia.

T. Y. Al-Naffouri is with the Electrical Engineering Department, King Abdullah University of Science and Technology (KAUST), Thuwal 23955, Saudi Arabia, and also with the Electrical Engineering Department, King Fahd University of Petroleum and Minerals (KFUPM), Dhahran 3126, Saudi Arabia (e-mail: Tarig.Ahmed@kaust.edu.sa; Tareq.Alnaffouri@kaust.edu.sa).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2014.2340818

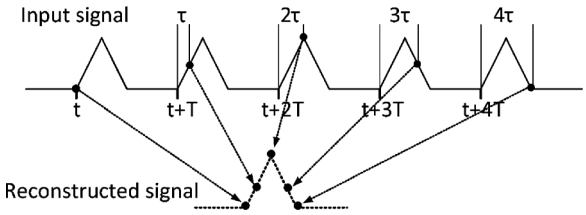


Fig. 1. Illustration of equivalent-time sampling.

allow for UWB signal sampling at the data *symbol rate*, which is much lower rate than the Nyquist frequency. It is obvious that the proposed configuration in [9] increases both cost and power consumption.

In this paper, we draw on the idea of *equivalent-time sampling* [10]–[12]. Equivalent-time sampling is a technique that has been widely used in digital oscilloscopes to capture a *repetitive signal* using a sampling rate that is lower than the signal's Nyquist rate. An illustrative example of this scheme is depicted in Fig. 1. In equivalent-time samplers, the ADC is triggered at progressively increasing time intervals (in Fig. 1, the time interval is increased by $\tau$ each time). Instead of acquiring samples in rapid succession, the ADC digitizes only one point from several occurrences of the input waveform and uses the samples to recreate the shape of the signal. As a result, the acquisition is not limited by the conversion rate of the ADC [11]. This equivalent-time sampling scheme obviously adds more complexity to the functionality of the ADC in addition to the stringent timing requirement.

Unlike the equivalent-time sampling scheme used in digital oscilloscopes, the scheme proposed herein does not require progressive sampling. In the proposed scheme, the repetitive signal is sampled at uniform intervals. To pick the required signal samples from the different repetitions, we rely on the relationship between the sampling rate and the inter-pulse interval, as will be explained in the sequel. It will be shown that certain constraints on the number of signal repetitions, the repetition time window and the sub-sampling rate can be utilized to allow for *perfect* reconstruction of the repetitive signal in the noise-free case. This enables us to use highly sub-sampled data to produce *channel impulse response* (CIR) estimates sampled at the Nyquist rate or higher.

This paper is organized as follows. Section II describes the signal models. In Section III, the proposed equivalent-time approach is explained. In Section IV, the effect of errors due to clock drift is considered and ways for mitigating it are discussed. The actual channel estimation solution is given in Section V assuming perfect synchronization. In Section VI, the analytical performance of the proposed channel estimation method is derived. The synchronization problem is discussed in Section VII. Simulation results are presented in Section VIII. The paper is concluded in Section IX.

### A. Notations

We use upper-case bold-face letters to denote matrices (e.g., $\boldsymbol{\Sigma}$) and lower-case bold-face letters to denote vectors (e.g., $\boldsymbol{\lambda}$). A lower-case letter with a subscript denotes an element of a vector (e.g., $\lambda_i$ is the $i$'th element of the vector $\boldsymbol{\lambda}$). The notations $(\cdot)^H$ and $\mathrm{tr}(\cdot)$ are used to denote the Hermitian transpose and the trace of a matrix, respectively. The operation denoted by $\mathrm{diag}(\cdot)$ returns the vector that contains the diagonal elements of a matrix argument; for a vector argument $\mathrm{diag}(\cdot)$ returns the diagonal matrix whose diagonal entries are the vector elements. The statistical expectation operation is denoted by $\mathrm{E}(\cdot)$ and the estimated value is denoted by $\hat{(\cdot)}$. The symbol '-- $\rightarrow$' is used to indicate that the expression to the right approximately replaces the expression to the left. The real part, imaginary part and magnitude of a complex number are denoted by, $\Re(\cdots)$ and $\Im(\cdot)$ and $|\cdot|$, respectively, Finally, $||\cdot||_2$ denotes the Euclidean norm in the case of a vector, or the 2-induced norm in the case of a matrix.

## II. Signal Models

A received UWB signal can be approximated as a linear combination of scaled and delayed versions of the transmitted signal [13], [14]. This leads to the well-known convolution relationship, which, for discrete-time signal, is given by

$$y[n] = \sum_{k=-\infty}^{+\infty} a[n-k]h(k) + v[n], \tag{1}$$

where $a[n]$ is the $n$'th sample of the transmitted signal, $y[n]$ is the corresponding received signal sample, $h[n]$ is the CIR $n$'th sample, and $v[n]$ is a of sample noise that is assumed to be *additive white Gaussian noise* (AWGN) with zero mean and variance $\sigma_v^2$. Here it is assumed that the signals are sampled at regular time intervals according to a sampling rate, $f_s$. Note that the model above ignores effects such as diffraction and dispersion, which result in frequency dependent distortions of the individual echoes [15].

Customarily, for time-limited signals, the process in (1) is more conveniently expressed in the matrix form

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{h} + \boldsymbol{v}, \tag{2}$$

where $\boldsymbol{y} \in \mathbb{R}^{M\times 1}$ is the received signal vector, $\boldsymbol{h} \in \mathbb{R}^{N\times 1}$ is the CIR, $\boldsymbol{v} \in \mathbb{R}^{M\times 1}$ is AWGN, and $\boldsymbol{A} \in \mathbb{R}^{M\times N}$ is referred to herein as the *transmission matrix*, which has the following structure:

$$\boldsymbol{A} = \begin{bmatrix} a_0 & a_{-1} & a_{-2} & \cdots & a_{-N+1} \\ a_1 & a_0 & a_{-1} & \cdots & a_{-N+2} \\ a_2 & a_1 & a_0 & \cdots & a_{-N+3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{M-1} & a_{M-2} & a_{M-3} & \cdots & a_0 \end{bmatrix}. \tag{3}$$

In the above model, it is assumed that any two adjacent elements in a row or a column are separated by a sampling interval $T_s = 1/f_s$.

In (2), the columns of the matrix $\boldsymbol{A}$ are shifted versions of the transmitted signal vector $\boldsymbol{a} \in \mathbb{R}^{M\times 1}$. The first column contains the signal $\boldsymbol{a}$ as it is. In the second column, the signal is shifted (down) by one element. In the third column, an additional shift by one element is introduced, and so on. Thus, the elements of $\boldsymbol{A}$ satisfy the property $A_{i,j} = A_{i+1,j+1}, \forall i \in \{0,\ldots,M-1\}, j \in \{0,\ldots,N-1\}$, which makes $\boldsymbol{A}$ a *Toeplitz matrix*.

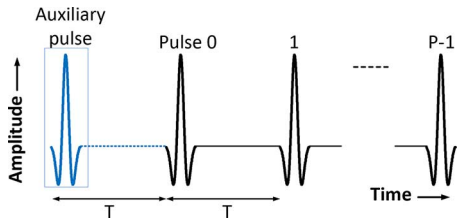


Fig. 2. The transmitted pulse train.

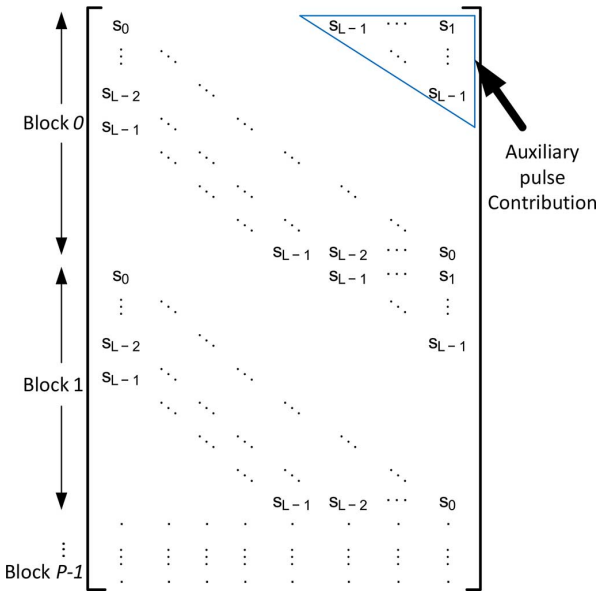


Fig. 3. Illustration of the structure of the transmission matrix $\boldsymbol{A}$. Each block corresponds to a sub-matrix $\boldsymbol{A}_b$.

The transmitted signal $\boldsymbol{a}$ consists of a priori known training sequence that is transmitted periodically, possibly as part of a data frame. Herein, we will ignore any transmitted data, and we will simply refer to $\boldsymbol{a}$ as the *transmitted signal* (in the channel estimation context). For the channel estimation method proposed in this paper, the training sequence is comprised of $P$ pulses that are transmitted sequentially at a regular time interval of $T$ seconds. An example of such a pulse train is shown in Fig. 2. The interval $T$ should be sufficiently large for the receiver to collect all the multipath arrivals pertaining to a certain transmitted pulse before the next transmission commences.

In this work, the transmission matrix $(\boldsymbol{A})$ for UWB channel estimation is constructed as follows. The number of columns of $\boldsymbol{A}$ is equal to $N$, which is chosen to coincide with the pulse interval $T$. Each column spans the duration of $P$ pulse intervals, leading to a row dimension of $M = PN$ for the matrix $\boldsymbol{A}$. The concatenation of the first row of $\boldsymbol{A}$ (starting from the last element) and the first column (starting from the first element) represents the samples of the transmitted signal $\boldsymbol{a}$ missing the first sample (see Fig. 3). Thus, $\boldsymbol{A}$ can be obtained directly from the vector $\boldsymbol{a}$.

To improve the transmission matrix structure, an *auxiliary* pulse is transmitted $T$ seconds prior to the onset of the actual transmission. This extra pulse can be viewed as the *past* data at the time point when the transmission of the actual pulse train starts (see Fig. 2). As a result of transmitting this auxiliary pulse, the elements of the augmented transmitted signal, $a_{-N+1}, \ldots, a_{-N+L-1}$, will become non-zero, or more specifically, these elements will be equal to the pulse samples $s_1, \ldots, s_{L-1}$. This choice gives $\boldsymbol{A}$ a *circulant* structure, i.e., the elements of $\boldsymbol{A}$ now satisfy the property $A_{i,j} = A_{i+1,j+1}, \forall i \in \{1, \ldots, M\}, j \in \{1, \ldots, N\}$; with $i+1$ and $j+1$ taken modulo $M = PN$ and $N$, respectively. Fig. 3 depicts an example of a transmission matrix with circulant structure. The figure shows that $\boldsymbol{A}$ is comprised of $P$ identical sub-matrix blocks. The benefit of the auxiliary pulse in giving a circulfant structure to the matrix $\boldsymbol{A}$ and its first sub-matrix is clear. Note that if the contribution of the auxiliary pulse is removed, the matrix $\boldsymbol{A}$ becomes a non-circulant Toeplitz matrix with one non-circulant sub-matrix at the top followed by $P-1$ circulant sub-matrices. As a result of the structure of the matrix $\boldsymbol{A}$, (2) can be written as

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_0 \\ \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_{P-1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{z} \\ \boldsymbol{z} \\ \vdots \\ \boldsymbol{z} \end{bmatrix} + \begin{bmatrix} \boldsymbol{v}_0 \\ \boldsymbol{v}_1 \\ \vdots \\ \boldsymbol{v}_{P-1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_b \\ \boldsymbol{A}_b \\ \vdots \\ \boldsymbol{A}_b \end{bmatrix} \boldsymbol{h} + \begin{bmatrix} \boldsymbol{v}_0 \\ \boldsymbol{v}_1 \\ \vdots \\ \boldsymbol{v}_{P-1} \end{bmatrix}, \tag{4}$$

where $\boldsymbol{A}_b \in \mathbb{R}^{N \times N}$; $\boldsymbol{y}_k, \boldsymbol{v}_k \in \mathbb{R}^{N \times 1}, \forall k \in \{0, \ldots, P-1\}$; and $\boldsymbol{z} = \boldsymbol{A}_b \boldsymbol{h}$. As explained above, $\boldsymbol{A}_b$ is a circulant matrix and it represents the transmission matrix when only one pulse is transmitted (in addition to an auxiliary pulse). Now, it can easily be seen that the model in (2) and (4) is a concatenation of $P$ circulant systems of the form

$$\boldsymbol{y}_k = \boldsymbol{z} + \boldsymbol{v}_k = \boldsymbol{A}_b \boldsymbol{h} + \boldsymbol{v}_k, \quad k = 0, \ldots, P-1. \tag{5}$$

Since the noise vectors $\boldsymbol{v}_k$ have identical statistics, $\boldsymbol{y}_k$ are statistically equivalent. Note that originally $N$ is the channel order. To get a unique solution in the single-pulse case with Nyquist sampling, at least $N$ observations are required. Therefore, the number of observations per pulse are chosen to be equal to $N$, and hence the total number of observations (excluding the auxiliary pulse) is equal to $M = PN$.

Finally, we note that the circulant property of the matrix $\boldsymbol{A}$ and it sub-matrices is exploited in two ways:

a) By making $\boldsymbol{A}$ circulant, all the sub-matrices of $\boldsymbol{A}$ are identical, which is essential for the proposed equivalent-time sampling method that will be presented in Section III.
b) The circulant property of the sub-matrix $\boldsymbol{A}_b$ is exploited for reducing the computational complexity of the channel estimation solution and its robust version in the presence of clock drift.

### A. Down-Sampled Signal Model

The model in (2) and (4) represents a *square* sampling case, where the sampling periods for the received signal and the CIR are identical. In the UWB case, as well as in some other practical situations, it is desirable to employ a cheaper ADC to sample the received signal at a lower rate than that of $\boldsymbol{h}$. Assuming a sampling period of $T_d > T_s$, the received signal becomes

$$\boldsymbol{y}_d = \boldsymbol{A}_d \boldsymbol{h} + \boldsymbol{v}_d, \tag{6}$$

where the subscript '$d$' indicates column down-sampling. Throughout this paper, we assume integer decimation ratio. For the channel estimation method presented in the next section, a decimation ratio equal to $P$ (the number of pulses) is required. Based on this decimation ratio, each column in (6) is obtained by re-sampling the corresponding column in (2) according to the new sampling rate $f_s/P$. Note that each column of $\boldsymbol{A}$ is composed of a repetitive pattern of $P$ cycles that are exactly identical. The sampling operation might pick some or all of the elements of the repetitive pattern depending of the values of $P$ and $N$. The resulting matrix $\boldsymbol{A}_d$ in (6) is of dimensions $N \times N$ (recall that $\boldsymbol{A}$ is $M \times N = PN \times N$). Also, the observation vector ($\boldsymbol{y}_d$) size is reduced to $N$ since it is obtained by down-sampling $\boldsymbol{y}$ (of size $M = PN$) $P$ times.

## III. UWB Channel Estimation Using Equivalent-Time Sampling

When the received signal $\boldsymbol{y}_d$ is sampled below the Nyquist rate, the individual pulses received from different paths will be under-represented according to the Nyquist sampling criterion. This leads to a loss of fidelity, that is, $\boldsymbol{y}_d$ no longer contains sufficient information to accurately estimate the CIR. In fact, for an extremely sub-sampled signal, some multipath components may be completely missed during the sampling process. For example, if the sampling interval $T_s = 1/f_s$ is larger than the pulse width, some multipath components will not be sampled even at a single time point, and hence cannot be recovered. To handel such challenging situations, an equivalent-time sampling approach is developed in this section.

To describe the proposed method for low-sampling rate channel estimation, consider the under-sampled system in (6) and the square-sampled system in (4). Without loss of generality, let us focus on the case where the higher sampling rate (the sampling rate required for the CIR) is the Nyquist rate. Assume that the received signal is sub-sampled by a factor $P$.

*Theorem 1:* Given a number of blocks $P$, for the two systems in (5) and (6) to be equivalent, a sufficient condition is that the two numbers $P$ and $N$ be co-prime.

*Proof:* Consider any column (vector) of the system in (4). Denote that column by $\boldsymbol{c}$. By down-sampling $\boldsymbol{c}$ by a factor $P$, we obtain the corresponding column in (6), which is denoted by $\boldsymbol{c}_d$. First, we want to prove that for co-prime $P$ and $N$, the elements of $\boldsymbol{c}_d$ are exactly the same as the elements of the corresponding column in (5), call it $\boldsymbol{c}_b$, taken in different order (or permutation).

The indices of the elements of $\boldsymbol{c}$ that are picked by the sampling process to form $\boldsymbol{c}_d$ are

$$\alpha_n = nP, \quad n = 0, 1, \ldots, N-1. \tag{7}$$

The actual value at each index, $\alpha_n$, is equal to the value of $\boldsymbol{c}_b$ at an index that is given by the mapping

$$\beta_n = \mathrm{mod}(\alpha_n, N), \quad n = 0, 1, \ldots, N-1; \tag{8}$$

where $\mathrm{mod}(\,\cdot, N)$ denotes the modulo-N operation.

Now, let us focus on two elements with indices $\alpha_i = iP$ and $\alpha_j = jP$, $0 \le i < j \le N-1$, whose mappings are $\beta_i = \mathrm{mod}\,(iP, N)$ and $\beta_j = \mathrm{mod}\,(jP, N)$, respectively. For these mappings the following relationships hold

$$\alpha_i = q_i N + \beta_i \tag{9}$$

$$\alpha_j = q_j N + \beta_j, \tag{10}$$

where $\{q_i, q_j\} \subset \{0, 1, 2, \ldots, P-1\}$ are the corresponding quotients. By subtracting (10) from (9), we obtain

$$\alpha_k = kP = q_k N + \beta_k, \tag{11}$$

where $k \stackrel{\mathrm{def}}{=} (i-j) \in \{1, 2 \ldots, N-2\}$; $q_k \stackrel{\mathrm{def}}{=} (q_i - q_j) \in \{0, 1, 2, \ldots, P-1\}$ and $\beta_k \stackrel{\mathrm{def}}{=} \beta_i - \beta_j$. It can easily be seen that $\alpha_k$ is another sampling index, the mapping of which is $\beta_k$.

Now, let us write $N = \mathcal{N}Z$ and $P = \mathcal{P}Z$; where $\mathcal{N}, \mathcal{P}$ and $Z$ are non-zero integers; $\mathcal{N}$ and $\mathcal{P}$ are co-prime. With this factorization, (11) yields

$$Z[k\mathcal{P} - q_k\mathcal{N}] = \beta_k. \tag{12}$$

From (12), it can be seen that $\beta_k = 0$ if and only if $k = \mathcal{N}$ and $q_k = \mathcal{P}$. Now, consider the case in which $P$ and $N$ are co-prime, i.e., $Z = 1$, $\mathcal{N} = N$ and $\mathcal{P} = P$. In this case, since $k \in \{1, 2 \ldots, N-2\}$ and $q_k \in \{0, 1, 2, \ldots, P-1\}$, then it is guaranteed that $k \neq \mathcal{N}$ and $q_k \neq \mathcal{P}$. Consequently,

$$\beta_k \neq 0 \Longrightarrow q_i - q_j \neq 0 \Longrightarrow q_i \neq q_j. \tag{13}$$

From (13), it follows that for co-prime $P$ and $N$, the operation in (8) is a *one-to-one* mapping. Hence,

$$\boldsymbol{c}_d = \mathrm{perm}(\boldsymbol{c}_b), \tag{14}$$

where $\mathrm{perm}(\,\cdot\,)$ is a permutation of the vector elements. From (14), we conclude that the system in (6) is identical to that in (5); the only difference is a row-wise permutation[1], which is ends of the proof of Theorem 1.

The implication of Theorem 1 is that we can utilize repetitions of the same data block for signal/paramter estimation while relaxing the sampling requirements. Specifically, the deficiency in sampling rate can be perfectly compensated for by exploiting extra blocks under the constraints suggested above.

## IV. Practical Considerations

In this section, practical effects are taken into account when applying the equivalent-time sampling method described in the previous section to the UWB channel estimation problem. Based on the conclusion of Section III, we can transmit a pulse train of $P+1$ pulses to obtain $P$ identical blocks of received signal that can be sampled at a rate of $f_N/P$ without loss of fidelity. Note that the extra pulse is required to make the system circulant.

Implicit in the above development are two assumptions:

a) The channel remains static throughout the measurement period (see (4)); and
b) there are no errors in transmission times.

[1]Row-wise permutations, applied to a linear system, do not affect the solution of that system.

For the first condition to be satisfied, the pulse train duration should be sufficiently small such that the channel variation from pulse interval to another is negligible. Throughout this paper, it will be assumed that the pulse train complies with this requirement. This condition can be easily satisfied in indoor environments since object movement speeds are sufficiently low (a few meters per second at most, while in Section VIII, we will see that the whole transmission interval is in the range of a few microseconds). In other words, the maximum achievable reduction in sampling rate is limited by how fast the environment changes.

The second condition is related to time synchronization; namely, the effect of time shifts in the transmitted pulse locations. The effect of mis-synchronized transmitted pulses is that the received blocks are not identical. One of the main causes of mis-synchronization is *clock drift*. Normally, clock drift is parameterized by the so-called *clock drift rate* that is given in *parts per million* (p.p.m.) [16], [17]. For a given drift rate $r$, the total drift at time $t$ seconds is given by

$$\delta_t = rt, \tag{15}$$

where $\delta_t$ is in microseconds if $r$ is in p.p.m.

Considering the models in (4) and (6), the direct result of clock drift is to create a mismatch between the received signals and their respective transmission matrices. In each case, the received signal is generated from a different transmission matrix, which is constructed from a transmission sequence where the pulse locations are shifted according to the drift rate. Let us assume that the auxiliary pulse transmission starts at a certain time instant. The next pulse (or the first pulse) will be shifted forward or backward depending on whether the drift is positive or negative. The same will happen to the following pulses, with increasing or decreasing shift levels. Fig. 4 depicts an example of a transmitted pulse train with positive clock drift effect. If the drift is reasonably small compared to the pulse width and the sampling rate, we can assume that the matrix that has actually produced the received signal is equal to the ideal transmission matrix $\boldsymbol{A}_d$ plus an error matrix $\boldsymbol{\Delta}_A$. Hence, the model in (6) is replaced by

$$\boldsymbol{y}_d = (\boldsymbol{A}_d + \boldsymbol{\Delta}_A)\boldsymbol{h} + \boldsymbol{v}_d, \tag{16}$$

where $\boldsymbol{\Delta}_A\boldsymbol{h}$ represents the contribution of the clock drift to the observation error. Note that the frequency distortion effect that has been alluded to above, can well be absorbed into the matrix $\boldsymbol{\Delta}_A$. However, for the sake of simplicity, we will consider clock drift as the sole contributor to the uncertainty represented by the matrix $\boldsymbol{\Delta}_A$.

Note that the ADC clock drift at the receiver can cause a similar effect to that caused by the transmitter's clock drift. For simplicity, we will ignore the effect of clock drif/jitter at the receiver. That effect can be modeled in a similar way to the transmitter's clock drift effect and it can also be absorbed into the error matrix $\boldsymbol{\Delta}_A$. The error matrix $\boldsymbol{\Delta}_A$ can be conceived as the difference between the two transmission matrices constructed from two different (drifted and non-drifted) pulse trains under a certain sampling rate. For a moderate drift rate, the support of $\boldsymbol{\Delta}_A$ will concentrate around the support locations of the matrix

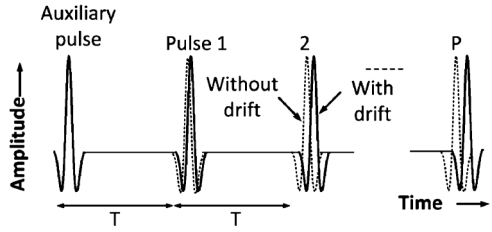


Fig. 4. An example of a transmitted pulse train with positive clock drift effect (continuous lines). The dotted lines represent the original sequence without clock drift effect.

$\boldsymbol{A}_d$. The model in (16) is the practical model that will be considered in the following developments.

In addition to the above two assumptions, it will also be assumed that the receiver has perfect knowledge of the time instant of the first arrival of the first pulse in the pulse train. Acquiring such synchronization information will be discussed in Section VII.

## V. Channel Estimation

Using the signal models (6) and (16), let us see how we can perform channel estimation. For the model in (6), i.e., when $\boldsymbol{\Delta}_A = \boldsymbol{0}$, the *least squares* (LS) estimator for the CIR is given by [18]

$$\hat{\boldsymbol{h}}_{\mathrm{LS}} = \left(\boldsymbol{A}_d^H \boldsymbol{A}_d\right)^{-1} \boldsymbol{A}_d^H \boldsymbol{y}_d, \tag{17}$$

For $\boldsymbol{\Delta}_A \neq 0$, (17) is expected to diverge depending on the drift rate. Note that for the AWGN case, (17) is also the *best linear unbiased estimator* (BLUE) [18].

Estimators that exploit the underlying model of the data are well-known to be more robust to different types of perturbations. These estimators include Bayesian estimators such as the *linear minimum mean squared error* (LMMSE) estimator, which, for zero-mean CIR and AWGN, is given by [18]

$$\hat{\boldsymbol{h}}_{\mathrm{LMMSE}} = \left(\boldsymbol{A}_d^H \boldsymbol{A}_d + \sigma_v^2 \boldsymbol{C}_{hh}^{-1}\right)^{-1} \boldsymbol{A}_d^H \boldsymbol{y}_d, \tag{18}$$

where $\boldsymbol{C}_{hh} = \mathrm{E}(\boldsymbol{h}\boldsymbol{h}^H)$ is the CIR covariance matrix. Note that the LMMSE approach assumes a stochastic CIR model contrary to the LS method that assumes a deterministic one. In reality, CIR models are stochastic, which gives a clear advantage to the LMMSE over the LS estimator. However, in practice, the channel model cannot be precisely known, and consequently the required statistics (i.e., $\boldsymbol{C}_{hh}$ and $\sigma_v^2$) are not available in most practical cases. A less strict version of the estimator in (19) can be obtained by assuming the CIR to be white (i.e., $\boldsymbol{C}_{hh} \dashrightarrow \sigma_h^2 \boldsymbol{I}$), which results in the *white* LMMSE (WLMMSE) estimator:

$$\hat{\boldsymbol{h}}_{\mathrm{WMMSE}} = \left(\boldsymbol{A}_d^H \boldsymbol{A}_d + \frac{\sigma_v^2}{\sigma_h^2}\boldsymbol{I}\right)^{-1} \boldsymbol{A}_d^H \boldsymbol{y}_d. \tag{19}$$

To deal with the *model uncertainties* (also known as *error-in-variables*) in linear systems, when no statistical information is available, a number of formalisms have been suggested. These include *total least squares* (TLS) based methods [19], [20], $H^\infty$ methods [21], and *bounded data uncertainty* (BDU) methods

[22]. The $H^\infty$ methods are unsuitable since they are derived for state-space models. In general, TLS is a good match for solving the problem in (16); however, the approach is reported to overestimate the uncertainty at times, thus leading to erroneous solutions. Such conservative behavior is avoided in the BDU approach since a bound on the uncertainty is imposed [22]. Because of this latter property, the BDU approach will be pursued in this paper. The formalism is also found to be well suited for the scenario under consideration and can be applied in a computationally efficient manner, as will be demonstrated in this section.

Based on [22], the CIR estimation problem in (16) can be formulated as a min-max problem. Namely, the solution is obtained by solving

$$\min_h \max \|(\boldsymbol{A}_d + \boldsymbol{\Delta}_A)h - (\boldsymbol{y} - \boldsymbol{v}_d)\|_2 \quad \text{subject to : } \|\boldsymbol{\Delta}_A\|_2 \le \eta, \|\boldsymbol{v}\|_2 \le \eta_v, \tag{20}$$

where $\eta$ is the upper bound on the 2-induced norm of $\boldsymbol{\Delta}_A$ and is assumed to be known, and $\eta_v$ is the upper bound on the Euclidian norm of $\boldsymbol{v}_d$ and will turn out to be irrelevant to the solution of (20). It can be proved that the min-max problem in (20) is equivalent to the minimization problem [22]

$$\min_h \|\boldsymbol{A}_d h - y\|_2 + \eta\|h\|_2 + \eta_v \tag{21}$$

whose solution takes the form [22]

$$\hat{\boldsymbol{h}}_{\text{BDU}} = (\boldsymbol{A}_d^H \boldsymbol{A}_d + \gamma \boldsymbol{I})^{-1} \boldsymbol{A}_d^H \boldsymbol{y}_d. \tag{22}$$

Here $\gamma$ is a regularization parameter that is obtained by solving the *secular equation* [22]

$$\boldsymbol{y}_d^H \boldsymbol{U}(\boldsymbol{\Sigma}^2 - \eta^2 \boldsymbol{I})(\boldsymbol{\Sigma}^2 + \gamma \boldsymbol{I})^{-2} \boldsymbol{U}^H \boldsymbol{y}_d = 0, \tag{23}$$

where $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\lambda})$, with $\boldsymbol{\lambda}$ being the vector of the singular values of $\boldsymbol{A}_d$; and $\boldsymbol{U} \in \mathbb{R}^{N \times N}$ is the matrix of eigenvectors of $\boldsymbol{A}_d \boldsymbol{A}_d^H$.

It is shown in [22] that for (22) to be the unique solution of (21), $\eta$ has to satisfy the inequality

$$\eta_l < \eta < \eta_u, \text{where} \quad \eta_l = \frac{\|\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_d\|_2}{\|\boldsymbol{\Sigma}^{-2}\boldsymbol{y}_d\|_2}, \quad \text{and} \quad \eta_u = \frac{\|\boldsymbol{A}_d^H \boldsymbol{y}_d\|_2}{\|\boldsymbol{y}_d\|_2}. \tag{24}$$

Note that for $\gamma = 0$, the solution in (22) coincides with the LS solution of (17).

In order to find the CIR estimator given in (22), a number of computations need to be performed. The following subsections will discuss how to carry out each of these computations in an efficient manner.

### A. Computing the Singular Values $\boldsymbol{\lambda}$

In a practical UWB system the matrix $\boldsymbol{A}_d$ is large and finding the singular values can be computationally demanding. To deal with this issue, we exploit the equivalence of the matrices $\boldsymbol{A}_d$ and $\boldsymbol{A}_b$, and the circulant property of the latter matrix. Namely, we have

$$\boldsymbol{A}_d^H \boldsymbol{A}_d = \boldsymbol{A}_b^H \boldsymbol{A}_b = \boldsymbol{B}. \tag{25}$$

The matrix $\boldsymbol{B}$ is a circulant matrix, and hence, its eigenvalues can be computed using the *fast Fourier transform* (FFT) of its first row, $\boldsymbol{b}_1$ [23]. From the relationship between the singular values of $\boldsymbol{A}_d$ and the eigenvalues of $\boldsymbol{B} = \boldsymbol{A}_d^H \boldsymbol{A}_d$, we obtain

$$\boldsymbol{\lambda} = \sqrt{\text{fft}\left(\boldsymbol{b}_1^H\right)}. \tag{26}$$

Note that the eigenvalue above can be pre-computed using any suitable method and stored to be used for computing the solution in (22).

### B. Finding $\gamma$

To find $\gamma$, (23) is iteratively solved. Since (23) is differentiable, the Newton's method [24] is one of the good options to carry out the task. Using the Newton's method, the iterations for finding $\gamma$, starting from an initial guess, $\gamma[0]$, are given by

$$\gamma[n+1] = \gamma[n] + \frac{\boldsymbol{y}_d^H \boldsymbol{U}(\boldsymbol{\Sigma}^2 - \eta^2 \boldsymbol{I})(\boldsymbol{\Sigma}^2 + \gamma[n]\boldsymbol{I})^{-2} \boldsymbol{U}^H \boldsymbol{y}_d}{2\boldsymbol{y}_d^H \boldsymbol{U}(\boldsymbol{\Sigma}^2 - \eta^2 \boldsymbol{I})(\boldsymbol{\Sigma}^2 + \gamma[n]\boldsymbol{I})^{-3} \boldsymbol{U}^H \boldsymbol{y}_d}. \tag{27}$$

The iterations in (27) reach convergence when $|\gamma[n+1] - \gamma[n]| < \epsilon$, where $\epsilon$ is a sufficiently small value. For (27) to converge to the only positive root, a judicious choice of the initial value $\gamma[0]$ is needed. Throughout this work, the initial value $\gamma[0] = 0$ is used. When tested in simulation, this initialization led to convergence after few iterations in more that 99% of the cases. Due to the involvement of mostly diagonal matrices, (27) can be evaluated with moderate computational cost.

### C. Computing the Inverse $(\boldsymbol{A}_d^H \boldsymbol{A}_d + \gamma \boldsymbol{I})^{-1}$

We can calculate the inverse $(\boldsymbol{A}_d^H \boldsymbol{A}_d + \gamma \boldsymbol{I})^{-1}$ that appears in the BDU solution (22) by again exploiting the circulant property of $\boldsymbol{A}_d^H \boldsymbol{A}_d$, which allows us to diagonalize it using the FFT matrix $\boldsymbol{F}$ as

$$\boldsymbol{A}_d^H \boldsymbol{A}_d = \boldsymbol{F}\boldsymbol{\Sigma}^{\boldsymbol{2}}\boldsymbol{F}^H, \tag{28}$$

where $\boldsymbol{\Sigma}^{\boldsymbol{2}} = \text{diag}(\boldsymbol{\lambda}^2)$. Hence, we can write

$$\left(\boldsymbol{A}_d^H \boldsymbol{A}_d + \gamma \boldsymbol{I}\right)^{-1} = \boldsymbol{F}(\boldsymbol{\Sigma}^{\boldsymbol{2}} + \gamma \boldsymbol{I})^{-1} \boldsymbol{F}^H. \tag{29}$$

This equation shows that the process of CIR estimation under the BDU model can be carried out without demanding much computational load. Note that while the matrix inverses required for evaluating the LS, LMMSE and WLMMSE solutions can all be pre-computed, the BDU solution requires re-computing the matrix inverse $(\boldsymbol{A}_d^H \boldsymbol{A}_d + \gamma \boldsymbol{I})^{-1}$ for each new observation vector ($\gamma$ is data dependent). To reduce the computational complexity involved, we capitalize on the circulant property of the data model. The solution of a circulant system can be implemented with a complexity of $O(N \log_2 N)$ compared to $O(N^2)$ for a Toeplitz system [23], which is a significant difference in computational complexity for a large value of $N$.

The method for CIR estimation with BDU can be summarized as follows:

1) Compute $\boldsymbol{\lambda}$, and then $\boldsymbol{\Sigma}$ using (26).
2) Find $\gamma$ using the iterations in (27).
3) Compute $\hat{\boldsymbol{h}}_{\mathrm{BDU}}$ based on (22) and (29) using the following steps:
   3.1 Multiply $\boldsymbol{y}$ by the matrix $\boldsymbol{A_d}$.
   3.2 Apply the FFT on the results.
   3.3 Divide (element-wise) by $\boldsymbol{\lambda}^2 + \gamma$.
   3.4 Finally, obtain $\hat{\boldsymbol{h}}_{\mathrm{BDU}}$ by applying the inverse FFT.

## VI. Analytical Performance

In the preceding section, we derived the BDU estimator of the CIR. This estimator minimizes the cost function in (21) (for a certain value of the $\eta$), which corresponds to solving the min-max problem (20). In most estimation problems, we are interested in minimizing the *mean squared error* (MSE), which is a different process that may not coincide with minimizing the maximum error as in (20). To get some insight into the MSE performance of the BDU solution, we conduct the following analysis. Starting from (22), the MSE is defined as

$$\begin{aligned}\mathrm{MSE} &= \mathrm{tr}\{\mathrm{E}[(\hat{\boldsymbol{h}}-\boldsymbol{h})(\hat{\boldsymbol{h}}-\boldsymbol{h})^H]\}\\ &= \mathrm{tr}[\mathrm{E}(\hat{\boldsymbol{h}}\hat{\boldsymbol{h}}^H)] - \mathrm{tr}[\mathrm{E}(\hat{\boldsymbol{h}}\boldsymbol{h}^H)]\\ &\quad - \mathrm{tr}[\mathrm{E}(\boldsymbol{h}\hat{\boldsymbol{h}}^H)] + \mathrm{tr}[\mathrm{E}(\boldsymbol{h}\boldsymbol{h}^H)],\end{aligned} \tag{30}$$

where $\boldsymbol{h}$ is the true CIR and, for simplicity, $\hat{\boldsymbol{h}}$ is used instead of $\hat{\boldsymbol{h}}_{\mathrm{BDU}}$. If we substitute for $\hat{\boldsymbol{h}}$ from (22) and manipulate we get the expression of the MSE in terms of the various system parameters. This derivation is detailed in Appendix A. The exact expression for the MSE can be simplified further by assuming that the CIR is *white*. This leads to the following expression of the MSE:

$$\begin{aligned}&\mathrm{MSE}_{\mathrm{white}}\\ &\quad= \sigma_v^2\mathrm{tr}\left[\left(\boldsymbol{A}_d^H\boldsymbol{A}_d+\bar{\gamma}\boldsymbol{I}\right)^{-1}\boldsymbol{A}_d^H\boldsymbol{A}\left(\boldsymbol{A}_d^H\boldsymbol{A}_d+\bar{\gamma}\boldsymbol{I}\right)^{-1}\right]\\ &\qquad+\sigma_h^2\mathrm{tr}\left[\left(\boldsymbol{A}_d^H\boldsymbol{A}_d+\bar{\gamma}\boldsymbol{I}\right)^{-1}\right.\\ &\qquad\times\left(\boldsymbol{A}^H\boldsymbol{\Delta}\boldsymbol{\Delta}^H\boldsymbol{A}-\bar{\gamma}\boldsymbol{A}^H\boldsymbol{\Delta}-\bar{\gamma}\boldsymbol{\Delta}^H\boldsymbol{A}+\bar{\gamma}^2\boldsymbol{I}\right)\\ &\qquad\left.\times\left(\boldsymbol{A}_d^H\boldsymbol{A}_d+\bar{\gamma}\boldsymbol{I}\right)^{-1}\right].\end{aligned} \tag{31}$$

To obtain an even simpler expression, we introduce several approximation (see Appendix A) to obtain the approximate MSE as

$$\begin{aligned}\mathrm{MSE}_{\mathrm{approx}} &\approx \sigma_v^2\sum_{i=0}^{N-1}\left(\frac{\lambda_i}{\lambda_i^2+\bar{\gamma}}\right)^2\\ &\quad+\sigma_h^2\sum_{i=0}^{N-1}\left\{\left[\frac{\Re(\omega_i)-\bar{\gamma}}{\lambda_i^2+\bar{\gamma}}\right]^2+\left[\frac{\Im(\omega_i)}{\lambda_i^2+\bar{\gamma}}\right]^2\right\},\end{aligned} \tag{32}$$

where $\boldsymbol{\omega}$ is a vector whose elements are the eigenvalues of $\boldsymbol{A}_d^H\boldsymbol{\Delta}_A$, and is approximated as:

$$\boldsymbol{\omega}\approx\mathrm{diag}[\boldsymbol{F}^H(\boldsymbol{A}^H\boldsymbol{\Delta}_A)\boldsymbol{F}]; \tag{33}$$

and $\bar{\gamma}$ is obtained from (27) by averaging over $\boldsymbol{y}_d$ and performing the iterations in a similar manner to (27). After manipulations, we obtain the following iterations:

$$\begin{aligned}\bar{\gamma}[n+1] &= \bar{\gamma}[n]\\ &\quad+\frac{\mathrm{tr}\left[\boldsymbol{U}(\boldsymbol{\Sigma}^2-\eta^2\boldsymbol{I})(\boldsymbol{\Sigma}^2+\bar{\gamma}[n]\boldsymbol{I})^{-2}\boldsymbol{U}^H\mathrm{E}\left(\boldsymbol{y}_d\boldsymbol{y}_d^H\right)\right]}{2\mathrm{tr}\left[\boldsymbol{U}(\boldsymbol{\Sigma}^2-\eta^2\boldsymbol{I})(\boldsymbol{\Sigma}^2+\bar{\gamma}[n]\boldsymbol{I})^{-3}\boldsymbol{U}^H\mathrm{E}\left(\boldsymbol{y}_d\boldsymbol{y}_d^H\right)\right]},\end{aligned} \tag{34}$$

where

$$\begin{aligned}\mathrm{E}\left(\boldsymbol{y}_d\boldsymbol{y}_d^H\right) &\dashrightarrow \sigma_v^2\boldsymbol{I}\\ &\quad+\sigma_h^2\left(\boldsymbol{A}_d\boldsymbol{A}_d^H+\boldsymbol{A}_d\boldsymbol{\Delta}^H+\boldsymbol{\Delta}_A\boldsymbol{A}_d^H+\boldsymbol{\Delta}_A\boldsymbol{\Delta}_A^H\right).\end{aligned} \tag{35}$$

The complete derivation of (31)–(34) is detailed in Appendix A. Now, contemplating the MSE in (32), we observe the following:

a) The MSE is composed of two main terms; the first term is the contribution of noise; and the second term is a function of the channel variance and drift.
b) Contrary to the first term, the second term is a bias term that does not asymptotically go to zero as the noise variance goes to zero.
c) In the special case where $\bar{\gamma}=0$ (i.e., in the LS case) and $\boldsymbol{\omega}=\boldsymbol{0}$ (i.e., no clock drif), $\mathrm{MSE}_{\mathrm{approx}}$ equals $N\sigma_v^2$, which is the performance of the LS estimator [18].
d) The benefit of the regularization parameter $\gamma$ in reducing the noise effect is clear. It also reduces the contribution of the *imaginary* parameter of the drift. However, the way the it is involved with the *real* parameter of the drift suggests that the regularization parameter, depending on its value, may produce a counter effect.

### A. Selecting $\eta$

First, recall that $\eta$ represents the upper limit on the *transmitted data uncertainty* as prescribed in (20). Being an upper limit means that there are infinitely many candidate values of $\eta$ that can be used. As the interest is customarily in the MSE performance, we can choose a value of $\eta$ that minimizes the MSE of the BDU estimator. This value can be obtained by differentiating the MSE expression, equating to zero and solving for $\bar{\gamma}$. The optimal value of $\eta$ can then be obtained by substituting $\bar{\gamma}$ in (23) and solving for $\eta$. This is a straightforward procedure, but unfortunately, the first derivative of the MSE expression (given in (31) or (32)) leads to an intractable order-$N$ nonlinear equation in $\bar{\gamma}$. To obtain a sub-optimal value of $\eta$, we optimize the MSE performance under zero clock-drift conditions. By substituting $\boldsymbol{\Delta}_A=\boldsymbol{0}$ in (32), differentiating and setting the derivative equal to zero, we obtain

$$\bar{\gamma}_{\mathrm{opt}}\approx\rho\stackrel{\mathrm{def}}{=}\frac{\sigma_v^2}{\sigma_h^2}, \tag{36}$$

By substituting (36) in (23) and solving for $\eta$, we obtain

$$\eta_{\mathrm{opt}}\approx\sqrt{\frac{N}{\sum_{i=0}^{N-1}\frac{1}{\lambda_i^2+\rho}}-\rho}. \tag{37}$$

From (36) and (37), it be can clearly seen that in the case where $\eta = \eta_{\text{opt}}$, the BDU estimator coincides with the WLMMSE estimator. In other words, by choosing $\eta$ according to (37), the BDU method performs exactly like a WLMMSE estimator. This result underscores the importance of exploiting the underlying model in providing robustness against uncertainties. However, this requires a priori knowledge that may not be available in practical situations. For example, in indoor environments, the statistical parameters of $\boldsymbol{h}$ can be time variant due to the movement of objects. Therefore, the value of the parameter $\rho$ is very difficult to estimate accurately in most realistic scenarios. To circumvent this hurdle, we resort to approximating $\eta_{\text{opt}}$ with its value at infinite SNR, i.e., when $\rho = 0$. This gives the $\eta$ value,

$$\eta_0 \approx \sqrt{\frac{N}{\sum_{i=0}^{N-1} \frac{1}{\lambda_i^2}}}, \tag{38}$$

which we use to obtain the final results presented in this paper. Note that $\eta_0$ is determined only by the transmission matrix and can, therefore, be calculated a priori. The rationale behind (38) is as follows:

1) Since we do not know the drift rate and the SNR, it is a good idea to optimize for the zero-noise zero-drift case.
2) It can be shown that $\eta_0$ increases with the increase in both noise and drift. However, the increase in $\eta_0$ is *insignificant* since the MSE becomes flatter when noise or drift increase, as will be shown in Section VIII. Therefore, calculating $\eta_0$ under zero-noise zero-drift conditions can be considered as a sub-optimal, yet reasonable, choice.

## VII. Synchronization

So far we have presented a low-sampling-rate solution to the UWB channel estimation problem. In the proceeding developments, we assumed perfect synchronization; that the receiver knows exactly the start of the received signal portion due to the $P$ transmitted pulses. In practice, however, some means of time acquisition and synchronization is required to gain such knowledge. The UWB synchronization problem has been discussed in a number of publications (see [1] and the references therein). Usually, a known sequence of modulated pulses is transmitted for this purpose. In this paper, we suggest utilizing the same channel estimation pulse sequence to achieve synchronization. We propose a simple synchronization method based on cross correlation.

In the proposed channel estimation scheme, the UWB receiver needs to acquire a data window of $N$ samples that starts exactly at the first multipath arrival pertaining to the first pulse. To achieve this, we need to estimate the first arrival time of the first pulse. Alternatively, we can detect the first arrival belonging to the auxiliary pulse and then calculate the first arrival time of the first pulse accordingly. The latter option is found more appealing; since a correlation based approach is used, it is helpful to incorporate the received data from all the pulses, including the auxiliary pulse, in the synchronization process.

Now, let us consider a received data window $w_d$ that starts exactly at the first arrival of the auxiliary pulse and that include all the arrivals attributed to the other $P$ pulses. The length of this window is $N_w = N + \lfloor N/P \rceil$, where $\lfloor N/P \rceil$ is the closest integer to $N/P$. Similar to (6), this data window is related to the transmitted pulse sequence by the relationship

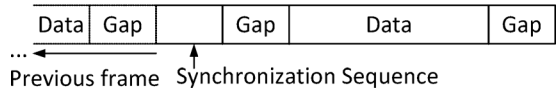


Fig. 5. Proposed data frame format.

$$\boldsymbol{w}_d = \boldsymbol{D}_d \boldsymbol{h} + \boldsymbol{v}_d, \tag{39}$$

where $\boldsymbol{D}_d \in \mathbb{R}^{N_w \times N}$ is an augmented sub-sampled transmission matrix. This matrix is obtained by down-sampling the matrix $\boldsymbol{D} = [\tilde{\boldsymbol{A}}_b^H, \boldsymbol{A}^H]^H$, where $\tilde{\boldsymbol{A}}_b$ is obtained from $\boldsymbol{A}_b$ by removing the contribution of the auxiliary pulse (see Fig. 3). This makes the matrix $\boldsymbol{D}$ a non-circulant matrix since the auxiliary pulse is not proceeded by any pulse transmission.

To detect the arrival of the pulse group (including the auxiliary pulse), we pick a test window $\boldsymbol{w}_i$ of length $N_w$. We define the correlation of this window as

$$\boldsymbol{\rho}_i = \boldsymbol{D}_d^H \boldsymbol{w}_i, \tag{40}$$

which represents the correlations of the transmitted pulse sequence (of $P + 1$ pulses) with $\boldsymbol{w}_i$ at all possible lags. For $\boldsymbol{\rho}_i$ to be useful in detecting the start of the pulse sequence, the window containing all the pulse sequence contributions ($\boldsymbol{w}_d$) has to produce higher correlation than any other window. This can be assured by proper design of the data transmission, for example, by using orthogonal codes, etc. To simplify the presentation of the proposed synchronization method, we pick a simple example where *pulse position modulation* (PPM) is used. We apply the proposed method to estimate the first time of arrival pertaining to the auxiliary pulse. To have good correlation properties, we suggest using a symbol rate that is lower than the synchronization pulse transmission rate. This makes any window that contains only data symbols produce less correlation values compared to a window that fully or partially contains synchronization/channel estimation pulse sequence. To provide more robustness, a guard interval with no transmission can be used before and after the synchronization pulse sequence. For example, the data frame format depicted in Fig. 5 can be used.

Based on the data frame format in Fig. 5, the proposed synchronization method looks for the synchronization point by searching over $N_f$ samples of data that correspond to one frame duration. We start by picking the first window $\boldsymbol{w}_0$ of $N_w$ samples. We calculate $\boldsymbol{\rho}_0$ using (40). Next, the window is slid right by one sample and the new correlation is calculated. The process is repeated until the last element of the $N_f$ sample vector is reached. For each window we also calculate the norm of the correlation $\|\boldsymbol{\rho}_i\|_2$. This scalar quantity increases as the test window passes through the synchronization sequence until a peak is reached. Then the value of $\|\boldsymbol{\rho}_i\|_2$ starts to drop as the test window passes away from the synchronization sequence (see Fig. 11). The point before which $\|\boldsymbol{\rho}_i\|_2$ starts to drop significantly is the start of the auxiliary pulse arrivals. To detect this point, first we calculate the normalized version of $\|\boldsymbol{\rho}_i\|_2$,

$$\mu_i = \|\boldsymbol{\rho}_i\|_2 / \max(\|\boldsymbol{\rho}_i\|_2), \quad \forall i = 0, \ldots, W-1, \tag{41}$$

where $W$ is the total number of windows. Note that the maximum value of $\mu_i$ is equal to 1.0. Next, we find the first index at which $\mu_i$ drops below a certain value $\zeta$. Let us denote this index by $n_1$. The synchronization index is calculated as $n_0 = n_1 - 1$, which is the estimated start of the auxiliary pulse reception. For channel estimation the receiver needs to pick the $N$ consecutive data samples starting from $n_0 + \lfloor N/P \rceil$.

### A. Synchronization Error

In this subsection, we consider the effect of synchronization error on the proposed channel estimation method. Let us denote the synchronization error in samples by $e$, which is obtained by subtracting the true synchronization point from the estimated one. Note these samples are based on the ADC rate not the Nyquist rate. Based on the value of $e$, we distinguish between the following four cases:

a) $e = 0$. In this case, channel estimation will produce an estimate of the CIR in which the first element corresponds to the first arrival.
b) $-\lfloor N/P \rceil \leq e < 0$. This coincides with an *early* synchronization point estimate. The corresponding CIR estimate is circularly shifted to the right (by $|e|$) relative to the true CIR.
c) $\lfloor N/P \rceil \geq e > 0$. This coincides with a *late* synchronization point estimate. The data window $\boldsymbol{w}_i$ is missing at least part of the first arrival data. The only way to obtain meaningful CIR estimate is to have an extra pulse at the end of the sequence to make up for the missing components. If this compensation is applied, the corresponding CIR estimate is circularly shifted to the left (by $|e|$) relative to the true CIR.
d) $|e| > \lfloor N/P \rceil$ channel estimation will not work in this case.

In both the second and third case, the receiver can detect the time of first arrival from the circularly shifted CIR and re-synchronize. This should be a straightforward process given the decaying nature of the channel profile and will not be considered in this paper. As will be proved by results, the occurrence of the fourth case is a very rare event.

## VIII. Simulations

### A. Channel Estimation

In order to test the proposed channel estimation method using equivalent-time sampling under different conditions, simulations were performed. The UWB channel models described in the IEEE 802.15.3a standard [6] were used. In all cases, the duration of the UWB pulse was 1 nanosecond. The target sampling rate was 4 GHz, which approximately coincides with the 10-dB Nyquist rate for the pulses that were used [25]. The inter-pulse interval was chosen to be the closest to 100 nanoseconds while satisfying the co-prime condition. For the 4-GHz sampling rate the inter-pulse interval is equivalent to 400 sampling periods. This is reduced to $N = 399$ to make it co-prime for all the values of the sub-sampling factor, $P = 5, 10, 20$ and 50, used in the simulations. All simulation results were averaged over $10^4$ simulation trials; each trial involved a different noise realization and a different CIR realization. In all these trials perfect synchronization was assumed.

Fig. 6(a) and (b) shows the variation of the MSE with $\eta$ when the pulse is a first derivative of a Gaussian function. Fig. 6(a) is obtained from simulation, while Fig. 6(b) is the approximate MSE given in (32). It can be seen that the analytical formulae provide a good approximation for the MSE. In spite of the deviations in some cases, the analytical formulae preserve the location of the optimal $\eta$ with sufficient accuracy. The same phenomenon is seen in Fig. 6(c) and (d), which are the counterparts of Fig. 6(a) and (b) for a second derivative of a Gaussian pulse. The effect of the SNR and drift is displayed in a more systematic pattern in the case of the second-derivative Gaussian pulse than in the case of the first-derivative pulse. In each individual figure, the vertical dotted line marks the location of $\eta_0$ calculated using (38). It can be seen that $\eta_0$ gives performance that is sufficiently close to the optimal one. As drift or noise increases, the minimum of the MSE shifts to the right (away from $\eta_0$), but at the same time, the MSE curves become flatter. This makes $\eta = \eta_0$ a good choice for the BDU solution since it remains close to the optimal value in most of the cases. It is evident from Fig. 6 that $\eta_0$ is different for different pulse shapes, which stresses the importance of (38) in choosing $\eta$. All the following results were obtained using $\eta = \eta_0$ for the second-derivative Gaussian pulse shape. The results are also based only on the CM1 model since no significant difference in performance was seen for the different (CM1, CM2, CM3 and CM4 [6]) models.

Fig. 7 depicts the MSE versus the SNR for zero drift and sub-sampling factor $P = 1, 5, 10, 20$ and 50. The performance of the BDU, LS, WLMMSE and LMMSE estimators is shown. The $P = 1$ case represents the Nyquist-sampling case. It can be seen that for each individual estimator, the performance for each sub-sampling factor is exactly equivalent to that obtained at Nyquist rate. The individual curves for the different $P$ values are actually indistinguishable. This emphasizes the ability of the proposed scheme to perfectly reconstruct the CIR from multiple sub-sampled observations. It can also be seen that the LS, WLMMSE and LMMSE estimators greatly outperform the BDU in the zero-drift case. This can be attributed to the model mismatch the BDU approach suffers due to the invalid assumption that $\boldsymbol{\Delta}_A \neq \mathbf{0}$. Subsequently, it will be shown that when clock drift is present, the BDU estimator can outperform the other three estimators.

Fig. 8 depicts the MSE versus the SNR for different sub-sampling factors, $P = 5, 10, 20$ and 50. The figure is plotted for a rather high drift rate of $r = 50$ p.p.m. (see [16], [17] for practical range of drift rates). To analyze the performance of the various estimators, we distinguish between to regimes: the low SNR and the high SNR regimes. From the figure, it is clear that the BDU estimator significantly outperforms the LS estimator in the low SNR regime. Both the BDU and the LS estimators do not use any statistical prior information. On the other hand, the WLMMSE estimator exhibits better performance than that of the BDU in the low SNR regime. However, The gap between the WLMMSE and the BDU estimators reduces as the SNR increases. The LMMSE achieves the best permeance at

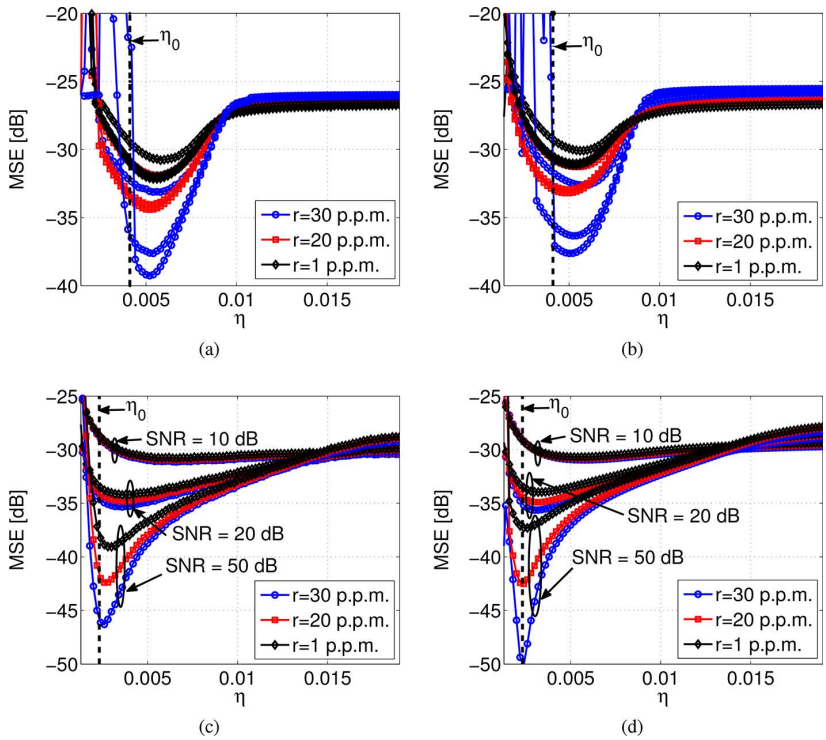


Fig. 6. MSE versus $\eta$: (a) first-derivative Gaussian pulse (simulation); (b) first-derivative Gaussian pulse (analytical); (c) second-derivative Gaussian pulse (simulation); (d) second-derivative Gaussian pulse (analytical).

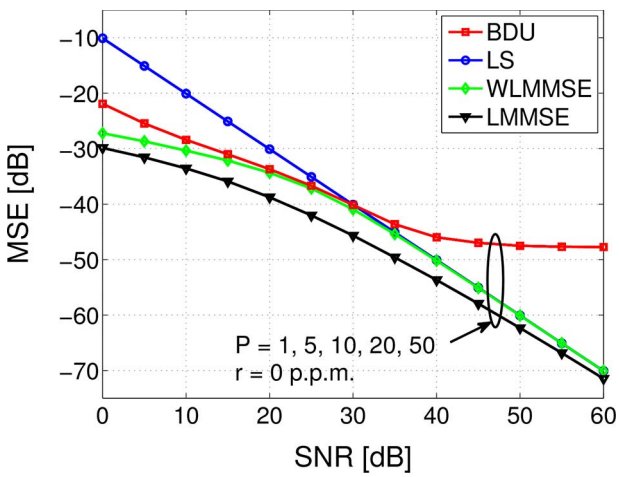


Fig. 7. MSE versus SNR for the BDU approach compared with the LS, WLMMSE and LMMSE approaches for zero clock drift and sub-sampling factor $P = 1, 5, 10, 20$ and 50.

low SNRs, which is attributed to the use of more prior information. As we move towards the high SNR regime, the performance of all the four estimators converges. For the $P = 20$ and $P = 50$ cases, it can be seen that the BDU performs better than all the other estimators. This can be explained by that for these (higher) sub-sampling rates, the total duration of the signal increases as more pulse cycles are incorporated. Consequently, the total drift (the drift rate multiplied by the total duration, see (15)) increases. This affect the other estimators more than the BDU estimator, which is more equipped to cope with such scenarios of data uncertainty.

In Fig. 9, $P$ is fixed to a value of 20, while $r$ is varied to take the values 5, 10, 50 and 80 p.p.m. The low SNR performance of the four estimators, relative to one another, is similar to that in Fig. 8; performance is greatly determined by the SNR in this regime. Again, the BDU estimator outperforms the other estimators at high SNRs and high drift rates, as in the cases of 50-p.p.m. and 80-p.p.m. drift rates.

From both Figs. 8 and 9, it can be concluded that in the low SNR regime, the performance of each of the four estimators is determined largely be the SNR. The effect of clock drift becomes more visible in the high SNR regime. In the presence of clock drift, the performance of each of the four estimators shows some sort of bias phenomenon, that performance tends to saturate towards high SNR. The BDU estimator shows some superiority at high SNRs and it outperforms the other three estimators when significant clock drift is present. This is further investigated in Fig. 10, which depicts performance against drift rate for $P = 20$ and the (high) SNR of 50 dB. It is evident from the figure that the gap between the BDU estimator and the other three estimators increases as the drift is increased by increasing the drift rate, or by incorporating more pulse cycles as for high values of $P$.

### B. Synchronization

To characterize the performance of the proposed synchronization method, the same simulation parameters described in the previous subsection were used. However, in these tests, a whole data frame was generated that contained the synchronization/channel estimation pulse sequence, as well as random data. The frame format depicted in Fig. 5 was used. The data was modulated using PPM with 4 symbols and data rate that was 1.5

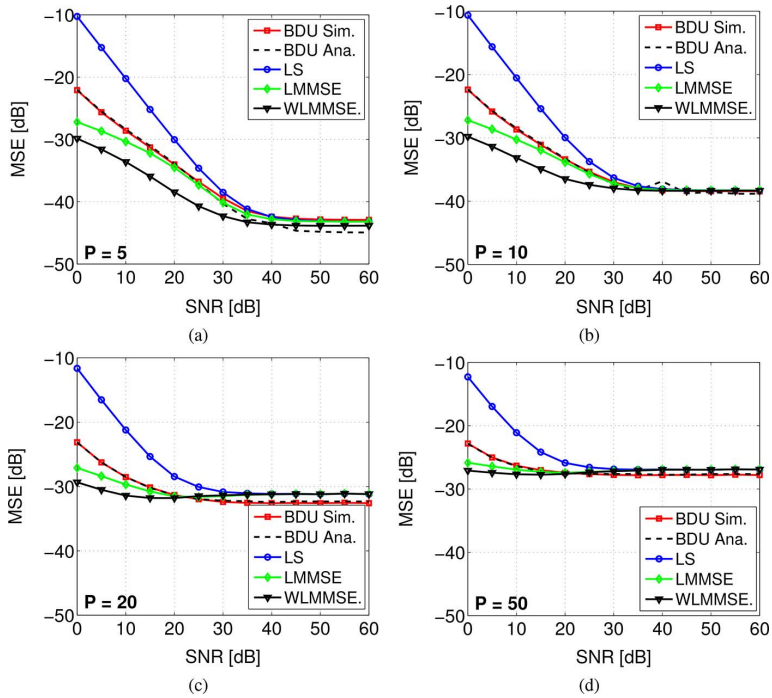


Fig. 8. MSE versus SNR for the BDU approach compared with the LS, WLMMSE and LMMSE approaches for a drift rate $r = 50$ p.p.m. and sub-sampling factor $P$ equal to (a) 5, (b) 10, (c) 20 and (d) 50.

times as high as the synchronization/channel estimation pulse rate. For simplicity, only 10 symbols were transmitted per frame and the simulation was confined to one frame proceeded by the gap and the data of another frame (as in Fig. 5). Hence, we had only one synchronization sequence in the setup. The search was carried out over all the simulated samples and was not confined only to $N_f$ samples, as recommended in Section VII. The gap interval was set equal to the duration of $P$ pulse cycles.

Fig. 11(a) (top) shows an example of the parameter $\mu_i$ calculated for all the simulated data windows. The bottom figure is the corresponding received signal. Fig. 11(b) is a zoom-in of Fig. 11(a) taken around the peak of $\mu_i$. It can be seen that $\mu_i$ exhibits a somewhat flat peak. This is explained as follows. Many of the data windows that start sufficiently close to the synchronization point and before it contain most of the energy of the pulse sequence (again because the profile is decaying). One would still expect this approximately flat peak to be increasing towards the right. However, due to the effect of noise the maximum (max) can occur any where on the flat peak and not necessarily on the far right. The significant drop in $\mu_i$ after the synchronization point is very clear. In our simulations, we used $\zeta = 0.995$ to detect the transition at $n_1$ and hence find the synchronization point $n_0$.

Fig. 12(a) and (b) plots the performance of the proposed synchronization method, which was derived from $10^4$ simulation trials. In Fig. 12(a), the *mean absolute error* (MAE), given in samples of the ADC frequency, is plotted against SNR for three different $P$ values and drift rates equal to 0 and 20 p.p.m. Note that the sampling periods are larger for larger values of $P$. The performance could have been presented in terms of Nyquist samples. However, it is found more convenient to present the results in terms of the ADC samples since that is the level of synchronization required in each case. In Fig. 12(a), the MAE is below 5 samples in most of the cases. Since the MAE can be dominated by outlier effect, a good performance indicator is to look at the percentage of cases in which the absolute value of the synchronization error ($|e|$) is equal to or less than a certain value. This is plotted in Fig. 12(b). It is shown that the percentage of cases in which perfect synchronization is achieved ($\%e = 0$) is approximately greater than or equal to 60%. The percent of cases in which $|e| \leq 1$ sample is always more that 80%. The percentage goes above 90% and approaches 100% for $|e| \leq 3$. This indicates a very good synchronization performance and the we can wither obtain the CIR estimate or its circularly shifted version, as explained in Section VII. Note that clock drift effect is more notable in the MAE and is almost unobservable when the distribution of the error is considered. This indicates an outlier effect when clock drift is present.

The plot of $\mu_i$ in Fig. 11 suggests that the synchronization process can be implemented in a more computationally efficient manner by using a two-stage procedure. In the first stage, a coarse search is performed by sliding the test window by a relatively large step instead of the proposed single-sample step. The step can be chosen such that at least one point is picked from the flat peak of $\mu_i$ as shown in Fig. 11. This will guarantee that the maximum of the coarse $\mu_i$ is one of the flat peak points. After finding the maximum of the coarse $\mu_i$, a fine search around the maximum can be carried out using a step of one sample. Using this fine search, the synchronization point can be located in the same way discussed in Section VII.

## IX. Conclusion

The problem of UWB channel estimation using sub-sampled observations was considered. A sub-sampling scheme using

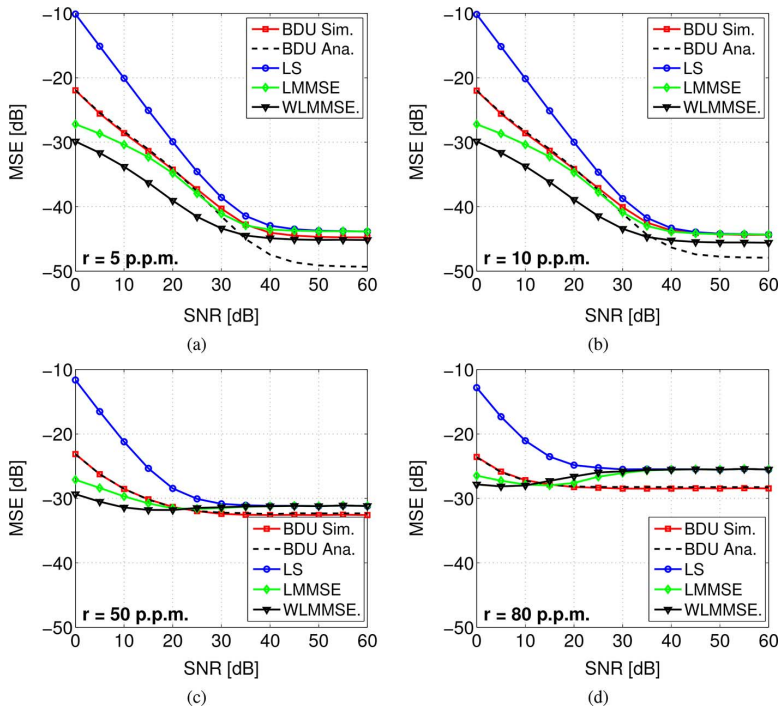


Fig. 9. MSE versus SNR for the BDU approach compared with the LS, WLMMSE and LMMSE approaches for a sub-sampling factor $P = 20$ and a drift rate $r$ equal to (a) 5, (b) 10, (c) 50 and (d) 80 p.p.m.

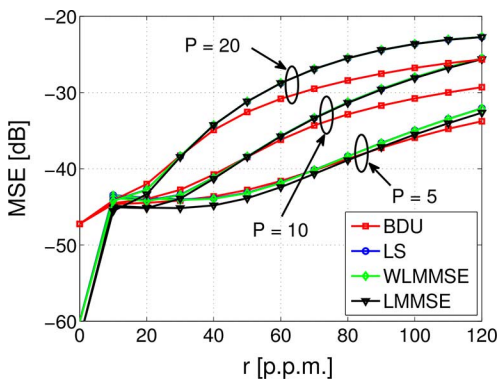


Fig. 10. MSE versus drift rate $r$ for the BDU approach compared with the LS, WLMMSE and LMMSE approaches for SNR $= 50$ dB and sub-sampling factor $P = 5, 10$ and 20.

multiple observations was proposed. The proposed scheme requires identical observations, and that the length of the observation window (in units of sampling periods) and the sub-sampling factor be co-prime. The effect of timing uncertainty due to clock drift was analyzed. An estimator based on the *bounded data uncertainty* (BDU) model was proposed to provide good channel estimation performance under different noise and clock drift conditions, without using statistical prior knowledge. When clock drift is present, the proposed estimator sufficiently outperforms the *least squares* (LS), the *linear minimum mean squared error* (LMMSE) and the *white linear minimum mean squared error* (WLMMSE) estimators in the high SNR case. At low SNRs, the proposed estimator performs remarkably better than the LS estimator that uses the same amount of information. In addition to channel estimation, the synchronization problem was also addressed. It was shown that the receiver can detect the first arrival of the transmitted pulse sequence with good accuracy. The effect of synchronization errors on channel estimation was also discussed.

The BDU approach proved to be capable of handling model errors due to clock drift. The same approach could be applied when the mismatch is due to pulse frequency distortion. Characterizing the performance of the BDU approach in such scenarios, and extending the approach if needed, constitute an interesting future work for the authors.

## Appendix A
## Derivation of the MSE

Starting from (30), we have four terms inside the trace. First let us look at the first term. Based on (22), this term can be written as

$$\mathrm{tr}[\mathrm{E}(\hat{\boldsymbol{h}}\hat{\boldsymbol{h}}^H)] = \mathrm{tr}\left[\left(\boldsymbol{A}_d^H\boldsymbol{A}_d + \bar{\gamma}\boldsymbol{I}\right)^{-1} \times \boldsymbol{A}_d^H \mathrm{E}\left(\boldsymbol{y}_d\boldsymbol{y}_d^H\right)\boldsymbol{A}_d\left(\boldsymbol{A}_d^H\boldsymbol{A}_d + \bar{\gamma}\boldsymbol{I}\right)^{-1}\right]. \quad \text{(A.1)}$$

Using (16), the expectation can be expanded as

$$\begin{aligned}\mathrm{E}\left(\boldsymbol{y}_d\boldsymbol{y}_d^H\right) &= \mathrm{E}\Big(\boldsymbol{A}_d\boldsymbol{h}\boldsymbol{h}^H\boldsymbol{A}_d^H + \boldsymbol{A}_d\boldsymbol{h}\boldsymbol{h}^H\boldsymbol{\Delta}^H \\ &\quad + \boldsymbol{\Delta}_A\boldsymbol{h}\boldsymbol{h}^H\boldsymbol{A}_d^H + \boldsymbol{\Delta}_A\boldsymbol{h}\boldsymbol{h}^H\boldsymbol{\Delta}_A^H + \boldsymbol{v_d}\boldsymbol{v}_d^H\Big) \\ &= \sigma_v^2\boldsymbol{I} + \boldsymbol{A}_d\boldsymbol{C}_{hh}\boldsymbol{A}_d^H + \boldsymbol{A}_d\boldsymbol{C}_{hh}\boldsymbol{\Delta}^H \\ &\quad + \boldsymbol{\Delta}_A\boldsymbol{C}_{hh}\boldsymbol{A}_d^H + \boldsymbol{\Delta}_A\boldsymbol{C}_{hh}\boldsymbol{\Delta}_A^H. \end{aligned} \quad \text{(A.2)}$$

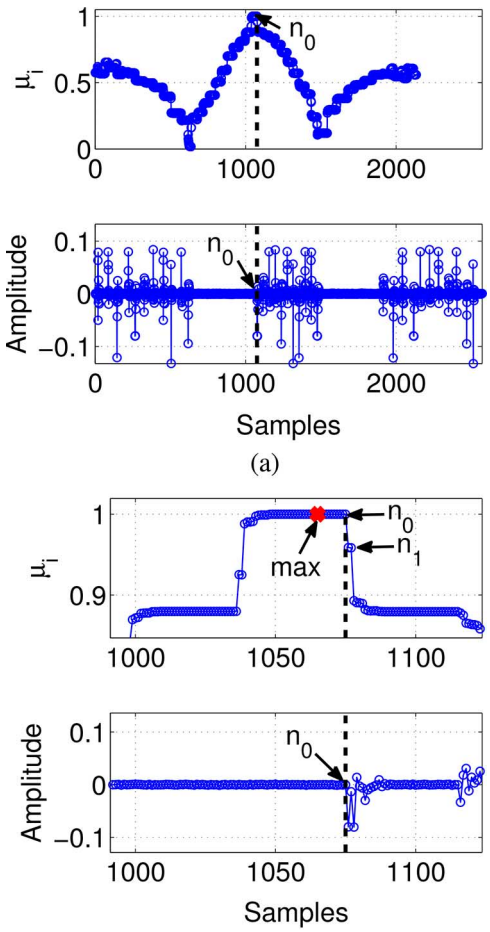


Fig. 11. An example plot of $\mu_i$ for $P = 10$ and 30-dB SNR: (a) full data, (b) a zoom-in around the peak.

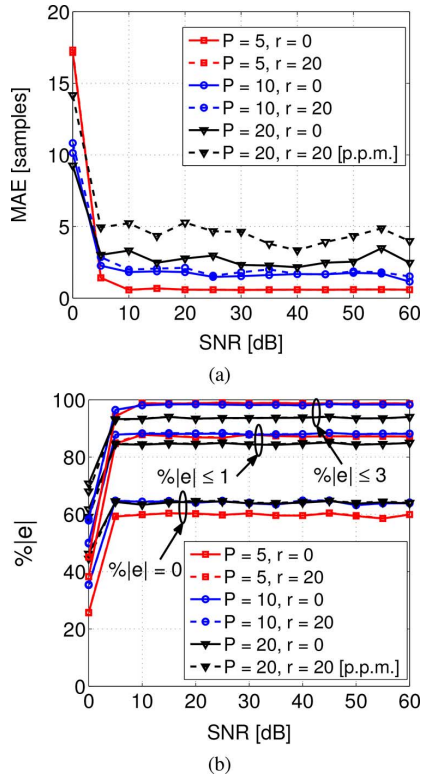


Fig. 12. Performance of the proposed synchronization method: (a) The mean absolute error (MAE) in samples; (b) Percentage of cases with absolute synchronization error less than or equal to selected values in samples.

The second term of the MSE is

$$\begin{aligned}
&\mathrm{tr}[\mathrm{E}(\hat{\boldsymbol{h}}\boldsymbol{h}^H)] \\
&\quad = \mathrm{tr}\left[\left(\boldsymbol{A}_d^H\boldsymbol{A}_d + \bar{\gamma}\boldsymbol{I}\right)^{-1}\boldsymbol{A}_d^H\mathrm{E}(\boldsymbol{y}_d\boldsymbol{h}^H)\right] \\
&\quad = \mathrm{tr}\left[\left(\boldsymbol{A}_d^H\boldsymbol{A}_d + \bar{\gamma}\boldsymbol{I}\right)^{-1}\boldsymbol{A}_d^H\right. \\
&\qquad \left.\times\ \mathrm{E}\left(\boldsymbol{A}_d\boldsymbol{h}\boldsymbol{h}^H + \boldsymbol{\Delta}_A\boldsymbol{h}\boldsymbol{h}^H + \boldsymbol{v}_d\boldsymbol{h}^H\right)\right] \\
&\quad = \mathrm{tr}\left[\left(\boldsymbol{A}_d^H\boldsymbol{A}_d + \bar{\gamma}\boldsymbol{I}\right)^{-1}\right. \\
&\qquad \left.\times\ \left(\boldsymbol{A}_d^H\boldsymbol{A}_d\boldsymbol{C}_{hh} + \boldsymbol{A}_d^H\boldsymbol{\Delta}_A\boldsymbol{C}_{hh}\right)\right], \qquad \text{(A.3)}
\end{aligned}$$

where it is assumed that $\mathrm{E}(\boldsymbol{v}_d\boldsymbol{h}^H) = 0$. Similarly,

$$\begin{aligned}
&\mathrm{tr}[\mathrm{E}(\boldsymbol{h}\hat{\boldsymbol{h}}^H)] \\
&\quad = \mathrm{tr}\left[\mathrm{E}\left(\boldsymbol{h}\boldsymbol{y}_d^H\right)\boldsymbol{A}_d\left(\boldsymbol{A}_d^H\boldsymbol{A}_d + \bar{\gamma}\boldsymbol{I}\right)^{-1}\right] \\
&\quad = \mathrm{tr}\left[\mathrm{E}\left(\boldsymbol{h}\boldsymbol{h}^H\boldsymbol{A}_d^H + \boldsymbol{h}\boldsymbol{h}^H\boldsymbol{\Delta}_A^H + \boldsymbol{h}\boldsymbol{v}_d^H\right)\right. \\
&\qquad \left.\times\ \boldsymbol{A}_d\left(\boldsymbol{A}_d^H\boldsymbol{A}_d + \bar{\gamma}\boldsymbol{I}\right)^{-1}\right] \\
&\quad = \mathrm{tr}\left[\left(\boldsymbol{C}_{hh}\boldsymbol{A}_d^H\boldsymbol{A}_d + \boldsymbol{C}_{hh}\boldsymbol{\Delta}_A^H\boldsymbol{A}_d\right)\right. \\
&\qquad \left.\times\ \left(\boldsymbol{A}_d^H\boldsymbol{A}_d + \bar{\gamma}\boldsymbol{I}\right)^{-1}\right]. \qquad \text{(A.4)}
\end{aligned}$$

Now, substituting (A.1)–(A.4) in (30), we obtain the expression for the *exact* MSE as

$$\begin{aligned}
&\mathrm{MSE}_{\mathrm{exact}} \\
&\quad = \sigma_v^2\mathrm{tr}\left[\left(\boldsymbol{A}_d^H\boldsymbol{A}_d + \bar{\gamma}\boldsymbol{I}\right)^{-1}\boldsymbol{A}_d^H\boldsymbol{A}\left(\boldsymbol{A}_d^H\boldsymbol{A}_d + \bar{\gamma}\boldsymbol{I}\right)^{-1}\right] \\
&\qquad + \mathrm{tr}\left[\left(\boldsymbol{A}_d^H\boldsymbol{A}_d + \bar{\gamma}\boldsymbol{I}\right)^{-1}\right. \\
&\qquad \times\ \left(\boldsymbol{A}^H\boldsymbol{\Delta}\boldsymbol{C}_{hh}\boldsymbol{\Delta}^H\boldsymbol{A} - \bar{\gamma}\boldsymbol{A}^H\boldsymbol{\Delta}\boldsymbol{C}_{hh}\right. \\
&\qquad \left.\left.-\bar{\gamma}\boldsymbol{C}_{hh}\boldsymbol{\Delta}^H\boldsymbol{A} + \bar{\gamma}^2\boldsymbol{C}_{hh}\right)\left(\boldsymbol{A}_d^H\boldsymbol{A}_d + \bar{\gamma}\boldsymbol{I}\right)^{-1}\right]. \qquad \text{(A.5)}
\end{aligned}$$

To obtain a simpler expression, we replace the CIR covariance by its *white* equivalent, $\boldsymbol{C}_{hh} \dashrightarrow \sigma_h^2\boldsymbol{I}$. This, after manipulation, results in the white MSE

$$\begin{aligned}
&\mathrm{MSE}_{\mathrm{white}} \\
&\quad = \sigma_v^2\mathrm{tr}\left[\left(\boldsymbol{A}_d^H\boldsymbol{A}_d + \bar{\gamma}\boldsymbol{I}\right)^{-1}\boldsymbol{A}_d^H\boldsymbol{A}\left(\boldsymbol{A}_d^H\boldsymbol{A}_d + \bar{\gamma}\boldsymbol{I}\right)^{-1}\right] \\
&\qquad + \sigma_h^2\mathrm{tr}\left[\left(\boldsymbol{A}_d^H\boldsymbol{A}_d + \bar{\gamma}\boldsymbol{I}\right)^{-1}\right. \\
&\qquad \times\ \left(\boldsymbol{A}^H\boldsymbol{\Delta}\boldsymbol{\Delta}^H\boldsymbol{A} - \bar{\gamma}\boldsymbol{A}^H\boldsymbol{\Delta} - \bar{\gamma}\boldsymbol{\Delta}^H\boldsymbol{A} + \bar{\gamma}^2\boldsymbol{I}\right) \\
&\qquad \left.\times\ \left(\boldsymbol{A}_d^H\boldsymbol{A}_d + \bar{\gamma}\boldsymbol{I}\right)^{-1}\right]. \qquad \text{(A.6)}
\end{aligned}$$

Further to the previous whiteness approximation, let us introduce the following replacements:

$$\begin{aligned}
\boldsymbol{A}_d^H\boldsymbol{\Delta}_A &\dashrightarrow \boldsymbol{F}\boldsymbol{\Omega}\boldsymbol{F}^H \\
\boldsymbol{\Delta}_A^H\boldsymbol{A}_d &\dashrightarrow \boldsymbol{F}\boldsymbol{\Omega}^H\boldsymbol{F}^H, \qquad \text{(A.7)}
\end{aligned}$$

where $\boldsymbol{\Omega}$ is obtained from

$$\boldsymbol{\Omega} = \mathrm{diag}\left[\mathrm{diag}\left(\boldsymbol{F}^H \boldsymbol{A}_d^H \boldsymbol{\Delta}_A \boldsymbol{F}\right)\right]. \quad \text{(A.8)}$$

Note that (A.8) implies that $\boldsymbol{A}_d^H \boldsymbol{\Delta}_A$ is approximated by a circulant matrix. The error matrix $\boldsymbol{\Delta}_A$ is the difference of the true transmission matrix $\dot{\boldsymbol{A}}_d$ and the matrix $\boldsymbol{A}_d$. The matrix $\dot{\boldsymbol{A}}_d$ is obtained by decimating a matrix $\dot{\boldsymbol{A}}$, which has a structure that is *approximately* similar to that of $\boldsymbol{A}$ (this is especially valid for small to moderate drift rate). As a result, the structure of $\boldsymbol{A}_d^H \boldsymbol{\Delta}_A = \boldsymbol{A}_d^H \dot{\boldsymbol{A}}_d - \boldsymbol{A}_d^H \boldsymbol{A}_d$ is approximately similar to that of $\boldsymbol{A}_d^H \boldsymbol{A}_d$, including the circulant property, which justifies (A.7) and (A.8).

Now, by substituting (A.7) together with (28) in (A.6) and manipulating based on the trace and FFT matrix properties, we obtain the *approximate* MSE

$$\begin{aligned}\mathrm{MSE}_{\mathrm{approx}} &= \sigma_v^2 \mathrm{tr}\left[\boldsymbol{\Sigma}^2 \left(\boldsymbol{\Sigma}^2 + \bar{\gamma}\boldsymbol{I}\right)^{-2}\right] \\ &\quad + \sigma_h^2 \mathrm{tr}\left\{\left(\boldsymbol{\Sigma}^2 + \bar{\gamma}\boldsymbol{I}\right)^{-2}\left[\bar{\gamma}^2 \boldsymbol{I} - 2\bar{\gamma}\Re(\boldsymbol{\Omega}) + |\boldsymbol{\Omega}|^2\right]\right\} \quad \text{(A.9)}\end{aligned}$$

Manipulating further yields (32).

Now, we revert to (34), which is derived from (27) by taking the expectation and the trace, i.e.,

$$\begin{aligned}\bar{\gamma}[n+1] &= \bar{\gamma}[n] \\ &+ \frac{\mathrm{tr}\left[\boldsymbol{U}(\boldsymbol{\Sigma}^2 - \eta^2\boldsymbol{I})(\boldsymbol{\Sigma}^2 + \bar{\gamma}[n]\boldsymbol{I})^{-2}\boldsymbol{U}^H \mathrm{E}(\boldsymbol{y}_d \boldsymbol{y}_d^H)\right]}{2\mathrm{tr}\left[\boldsymbol{U}(\boldsymbol{\Sigma}^2 - \eta^2\boldsymbol{I})(\boldsymbol{\Sigma}^2 + \bar{\gamma}[n]\boldsymbol{I})^{-3}\boldsymbol{U}^H \mathrm{E}(\boldsymbol{y}_d \boldsymbol{y}_d^H)\right]}, \quad \text{(A.10)}\end{aligned}$$

where $\mathrm{E}(\boldsymbol{y}_d \boldsymbol{y}_d^H)$ is substituted from (A.2) for the exact MSE. For the white MSE, we have

$$\begin{aligned}\mathrm{E}\left(\boldsymbol{y}_d \boldsymbol{y}_d^H\right) &= \sigma_v^2 \boldsymbol{I} \\ &\quad + \sigma_h^2\left(\boldsymbol{A}_d \boldsymbol{A}_d^H + \boldsymbol{A}_d \boldsymbol{\Delta}^H + \boldsymbol{\Delta}_A \boldsymbol{A}_d^H + \boldsymbol{\Delta}_A \boldsymbol{\Delta}_A^H\right). \quad \text{(A.11)}\end{aligned}$$

## REFERENCES


[1] G. B. Giannakis and L. Yang, "Ultra-wideband communications: An idea whose time has come," *IEEE Signal Process. Mag.*, vol. 21, no. 6, pp. 26–54, Nov. 2004.

[2] J. L. Paredes, G. R. Arce, and Z. Wang, "Ultra-wideband compressed sensing: Channel estimation," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 3, pp. 383–395, Oct. 2007.

[3] Y. Vanderperren, W. Dehaene, and G. Leus, "Performance analysis of a flexible subsampling receiver for pulsed uwb signals," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4134–4142, Aug. 2009.

[4] V. Lottici, A. D"Andrea, and U. Mengali, "Channel estimation for ultra-wideband communications," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 9, pp. 1638–1645, Dec. 2002.

[5] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[6] J. Foerster *et al.*, "Channel modeling sub-committee report final," *IEEE P802.15 Wireless Pers. Area Netw. P802.15-02/490r1-SG3a*, Feb. 2003.

[7] I. Maravić, J. Kusuma, and M. Vetterli, "Low-sampling rate UWB channel characterization and synchronization," *J. Commun. Netw.*, vol. 5, no. 2, pp. 319–327, 2003.

[8] Z. Wang and X. Yang, "Ultra wide-band communications with blind channel estimation based on first-order statistics," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2004, vol. 4, pp. 529–532.

[9] Y. (Geoffrey) Li, A. F. Molisch, and J. Zhang, "Channel estimation and signal detection for UWB," presented at the IEEE WPMC, Yokosuka, Japan, 2003.

[10] R. A. Lawton, S. M. Riad, and J. R. Andrews, "Pulse and time-domain measurements," *Proc. IEEE*, vol. 74, no. 1, pp. 77–81, Jan. 1986.

[11] K. Lawton and E. McConnell, National Instruments Corporation, "Equivalent Time Sampling for High-Speed Repetitive Signals Using E Series Boards and NIDAQ Software, Application Note 066," 1995.

[12] E. Moreno-Garcia, J. M. De la Rosa-Vazquez, and O. Alonzo-Larraga, "An approach to the equivalent-time sampling technique for pulse transient measurements," in *Proc. 16th Int. Conf. Electron., Commun. Comput. (CONIELECOMP)*, 2006, pp. 34–34.

[13] S. Gezici, "Theoretical limits for estimation of periodic movements in pulse-based UWB systems," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 3, pp. 405–417, Oct. 2007.

[14] Y. Zhou, Y. L. Guan, and C. L. Law, "Modified phase-only correlator with kurtosis-based amplified-noise suppression," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3341–3345, Nov. 2010.

[15] F. Molisch, K. Balakrishnan, C. Chin Chong, S. Emami, A. Fort, J. Karedal, J. Kunisch, H. Schantz, U. Schuster, and K. Siwiak, "IEEE 802.15.4a channel model—Final report," 2004 [Online]. Available: http://www.ieee802.org/15/pub/TG4a.html

[16] B. Denis, J.-B. Pierrot, and C. Abou-Rjeily, "Joint distributed synchronization and positioning in UWB ad hoc networks using TOA," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 4, pp. 1896–1911, Jun. 2006.

[17] A. D'Amico, L. Taponecco, and U. Mengali, "Ultra-wideband TOA estimation in the presence of clock frequency offset," *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, pp. 1606–1616, Apr. 2013.

[18] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[19] S. Huffel, J.-B. Pierrot, and C. Abou-Rjeily, *The Total Least Squares Problem: Computational Aspects and Analysis*. Philadelphia, PA, USA: SIAM, 1991.

[20] I. Markovsky and S. Van Huffel, "Overview of total least-squares methods," *Signal Process.*, vol. 87, no. 10, pp. 2283–2302, 2007.

[21] B. Hassibi, A. Sayed, and T. Kailath, *Indefinite Quadratic Estimation and Control: A Unified Approach to $H^2$ and $H^\infty$ Theories*. Philadelphia, PA, USA: SIAM, 1999.

[22] S. Chandrasekaran, G. Golub, M. Gu, and A. Sayed, "Parameter estimation in the presence of bounded data uncertainties," *SIAM J. Matrix Anal. Appl.*, pp. 235–252, 1998.

[23] M. Chen, "On the solution of circulant linear systems," *SIAM J. Numer. Anal.*, vol. 24, no. 3, pp. 668–683, Jun. 1987.

[24] C. Zarowski, *An Introduction to Numerical Analysis for Electrical and Computer Engineers*. New York, NY, USA: Wiley, 2004.

[25] M. Benedetto, T. Kaiser, A. Molisch, I. Oppermann, C. Politano, and D. Porcino, *UWB Communication Systems—A Comprehensive Overview*. New York, NY, USA: Hindawi, 2006.



**Tarig Ballal** received the Ph.D. degree from the School of Computer Science and Informatics, University College Dublin, Ireland, in 2011, an M.Sc. degree in telecommunications from Blekinge Institute of Technology, Sweden, in 2005, and a B.Sc. degree in electrical engineering from University of Khartoum, Sudan, in 2001.

He is a postdoctoral fellow in the Electrical Engineering Department, King Abdullah University of Science and technology (KAUST), Saudi Arabia, since September 2012. His current research focuses on acoustic and electromagnetic movement sensing, tracking and localization, with particular focus on low-cost and low-complexity methods. He is also interested in GNSS positioning and attitude determination, in addition to signal processing applications in wireless communications and biomedical sensing. Prior the joining KAUST, in the period from April 2011 to July 2012, he was a research engineer with the UCD Electrical Engineering Department and BiancaMed Ltd., Dublin, Ireland.

**Tareq Y. Al-Naffouri** (M'10) received the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 2004.

He was a visiting scholar at the California Institute of Technology, Pasadena, in 2005 and 2006, and a Fulbright scholar at the University of Southern California in 2008. He is currently an Associate Professor with the Electrical Engineering Department, King Fahd University of Petroleum and Minerals, and jointly at the Electrical Engineering Department, King Abdullah University of Science and Technology (KAUST) both in Saudi Arabia. His research interests lie in the areas of adaptive, statistical, and sparse signal processing and their applications to wireless communications, and in multiuser information theory.

Dr. Al-Naffouri has more than 90 publications in Journal and conference proceedings, nine standard contributions, five issued patents, and five pending.