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

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

Electrical Engineering Department, King Abdullah University of Science and Technology (KAUST),

Thuwal 23955-6900, Kingdom of Saudi Arabia.

Ph: + (966) 2 808 0298 , Fax: +(966) 2 802 0143.

E-mail: {Tarig.Ahmed, Tareq.Alnaffouri }@kaust.edu.sa.

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 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.

Key Words—ultra-wideband, UWB, channel estimation, sub-sampling, ADC, equivalent-time sampling, 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, 2, 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 parameter 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 CM_1 model (based on line-of-sight channel measurements at 0–4 m) has a cluster arrival rate of 0.0233 ns^{-1} and ray arrival rate of 2.5 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 $f_s \geq (107/0.1)/(43 \times 10^{-9}) \approx 25$ 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 sampling hardware on the other hand (the latter is already high for UWB). 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 was 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 devices. Improving this technique is practically limited by 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 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 was used to allow for sampling of UWB signals at the data*symbol rate* (data symbols are transmitted at a much lower rate compared to 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, 11, 12]. Equivalent-time sampling is a technique that has been used widely in digital oscilloscopes to capture a *repetitive signal* using a sampling rate which 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 τ 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 required.

Unlike the equivalent-time sampling scheme used in the digital oscilloscope, the scheme proposed herein does not require progressive sampling. In the proposed scheme, the repetitive signal is sub-sampled at uniform intervals. To pick the required signal samples from the different repetitions, we rely on the relationship between the sub-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 repeating 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.

The remainder of this paper is organized as follows. Section II describes the signal model required for the development of the proposed approach. In Section III, the proposed equivalent-time approach is derived assuming perfect time synchronization. In Section IV, the effect of synchronization error is considered and ways for mitigating it are discussed. The actual channel estimation solution is given in Section V. Simulation results are discussed in Section VII. The paper is concluded in Section VIII.

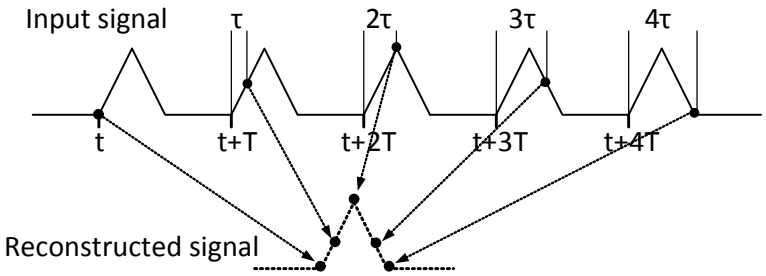

Fig. 1: Illustration of equivalent-time sampling.

I.I. Notations

We use upper-case bold-face letters to denote matrices (e.g., $\mathbf{\Sigma}$) 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., λ_i is the i'th element of the vector $\boldsymbol{\lambda}$). The notations $(.)^H$ and tr$(.)$ are used to denote the Hermitian transpose and the trace of a matrix, respectively. The operation denoted by diag(.) returns the vector that contains the diagonal elements of a matrix argument; for a vector argument diag(.) returns the diagonal matrix whose diagonal entries are the vector elements.The statistical expectation operation is denoted by E(.) and the estimated value is denoted by $\hat{(.)}$. The symbol '$\dashrightarrow$' 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, $\Re(.)$ and $\Im(.)$ and $|.|$, respectively, Finally, $||.||_2$ 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 the discrete-time case, 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, and $v[n]$ is a sample noise that is assumed to be *additive white Gaussian* (AWGN) with zero mean and variance σ_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, $\boldsymbol{h} \in \mathbb{R}^{N\times 1}$ is the CIR, $\boldsymbol{v} \in \mathbb{R}^{M\times 1}$ is additive noise, 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 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 $\boldsymbol{a} \in \mathbb{R}^{M+N-1\times 1}$. 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 $A_{i,j} = A_{i+1,j+1}, \forall i \in \{1, \cdots, M\}, j \in \{1, \cdots, N\}$ making $\boldsymbol{A}$ a *Toeplitz matrix*.

The transmitted signal $\boldsymbol{a}$ consists of an a priori known training sequence of data symbols that are transmitted periodically, possibly as part of data frames. Herein, we will simply 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, a training sequence is comprised of P pulses that are transmitted sequentially at a regular time interval, 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 transmitted signal $\boldsymbol{a}$, $a_{-N+1}, ..., a_{-N+L-1}$, will become non-zero, or more specifically, these elements will be equal to the pulse samples $s_1, ..., s_{L-1}$. This choice gives $\boldsymbol{A}$ a *circulant* structure, i.e., the elements of $\boldsymbol{A}$ now satisfy the property $A_{i,j} = A_{i+1,j+1}, \forall i \in \{1, \cdots, M\}, j \in \{1, \cdots, N\}$; with $i+1$ and $j+1$ taken modulo M and N, respectively. Fig. 3 depicts an example of a transmission matrix with circulant structure. The figure shows that $\boldsymbol{A}$ is comprized of P identical sub-matrix blocks. The benefit of the auxiliary pulse in making the matrix circulant is clear. Note that if the contribution of the auxiliary pulse is removed, the $\boldsymbol{A}$ 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 $\boldsymbol{A}$, Equation. (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 $\boldsymbol{A}_b \in \mathbb{R}^{N \times N}$; $\boldsymbol{y}_k, \boldsymbol{v}_k \in \mathbb{R}^{N \times 1}, \forall k \in \{0, \cdots, P-1\}$; 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 tranmsitted (in addition to the 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

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

Since the noise vectors $\boldsymbol{v}_k$ have identical statistics, $\boldsymbol{y}_k$ are statistically equivalent.

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

(a) By making $\boldsymbol{A}$ circulant, all the sub-matrices of $\boldsymbol{A}$ are identical, which is essential for the proposed

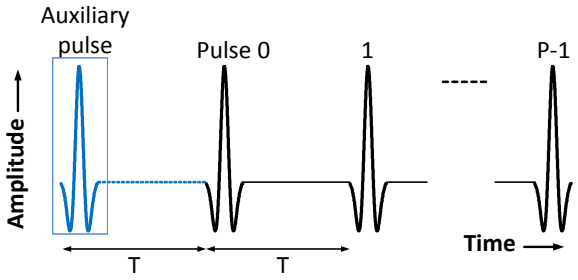

Fig. 2: The transmitted pulse train.

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 (see Section V further ahead).

II.I. Down-Sampled Signal Model

The model in Equation. (2) and Equation (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 utilize a cheaper ADC to sample the received signal at a lower rate than the one required for $\boldsymbol{h}$. Assuming a sampling period of $T_d = DT_s$, the received signal becomes

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

where the subscript 'd' indicates the down-sampling operation of the columns. Throughout this paper, we assume integer *decimation* ratio, i.e., $D \in \mathbb{N}$. Each column in (6) is obtained by re-sampling the corresponding column in (2) according to the new sampling rate f_s/D . 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 D and N.

The system in (6) has the same column dimension as the system in (2) and (4), and a row dimension equal to N/D (or the nearest integer less than N/D if N is not divisible by D).

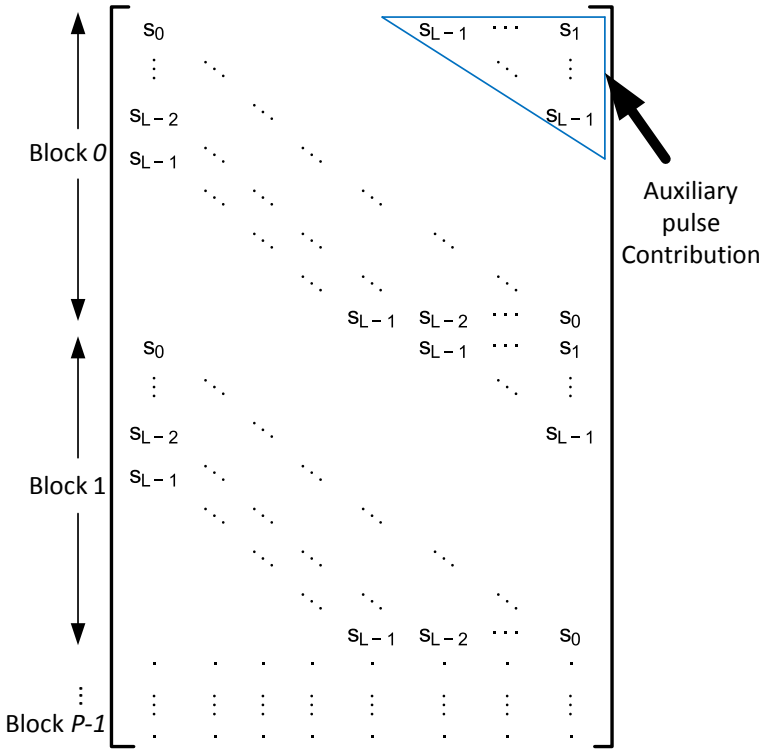

Fig. 3: Illustration of the structure of the transmission matrix $\boldsymbol{A}$. Each block corresponds to a sub-matrix $\boldsymbol{A}_b$.

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 $\boldsymbol{y}_d$, some multipath components may completely be missed by 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 in a straightforward manner. 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 output CIR) is the Nyquist rate. Assume that the received signal is sub-sampled by a factor D .

Theorem 1: *Given a number of blocks $D = P$, for the two systems in (5) and (6) to be equivalent, a*

sufficient condition is that the two numbers D 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 D, we obtain the corresponding column in (6). Denote the latter column by $\boldsymbol{c}_d$. We want to prove that for co-prime D 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 the $\boldsymbol{c}$ that are picked by the sampling process to form $\boldsymbol{c}_d$ are

$$\alpha_n = nD, n = 0, 1, \ldots, (N-1)D. \tag{7}$$

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

$$\beta_n = \text{mod}(\alpha_n, N), n = 0, 1, \ldots, (N-1)D, \tag{8}$$

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

Now, let us focus on two elements that correspond to $\alpha_i = iD$ and $\alpha_j = jD$, $0 \leq i < j \leq N-1$, whose mappings are $\beta_i = \mod(iD, N)$ and $\beta_j = \mod(jD, N)\}$, 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 $\{q_i, q_j\} \subset \{0, 1, 2, \ldots, D-1\}$ are the corresponding quotients. Now, consider $k \stackrel{def}{=} (i-j) \in \{1, 2 \ldots, N-2\}$. By subtracting (10) from (9), we obtain

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

where $q_k \stackrel{def}{=} (q_i - q_j) \in \{0, 1, 2, \ldots, D-1\}$ and $\beta_k \stackrel{def}{=} \beta_i - \beta_j$. It can easily be seen that α_k is another sampling index, the mapping of which is β_k.

Now, let us write $N = N_b Z$ and $D = D_b Z$, where $\{N_b, D_b, Z\} \subset \mathbb{N}$ with N_b and D_b being co-prime. Applying this factorization to Equation (11) yields

$$Z\left[kD_b - q_k N_b\right] = \beta_k. \tag{12}$$

From (12), it can be seen that $\beta_k = 0$ if and only if $k = N_b$ and $q_k = D_b$. Now, consider the case in which D and N are co-prime, i.e., $Z = 1$, $N_b = N$ and $D_b = D$. In this case, since $k \in \{1, 2 \ldots, N-2\}$ and $q_k \in \{0, 1, 2, \ldots, D-1\}$, then it is guaranteed that $k \neq N_b$ and $q_k \neq D_b$. 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 D and N, the operation in (8) is a *one-to-one* mapping. Therefore, we can write

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

where perm(.) is the permeation operation 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 the end 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 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 the Section III, we can transmit a pulse train of $D + 1$ pulses to obtain D identical blocks of received signal that can be sampled at a rate of f_N/D without loss of fidelity. Note that the extra pulse is required to make the system circulant.

Implicit in the above development are two assumptions for equivalent-time sampling to work:

(a) The channel remains *static* throughout the measurement period (see Equation (4)); and

(b) there are no errors in transmission times.

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 trains used herein comply with this requirement. This condition can be satisfied

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

easily in practice since object movement speeds (a few meters per second at most) are sufficiently low, in indoor environments, to consider the environment static during the whole transmission intervals (a few microseconds at most for the maximum number of pulses used in the simulations in Section VII). In other word, the maximum achievable reduction in sampling rate for the proposed method is limited by how fast the environment change.

The second condition is related to time synchronization; namely, the effect of any 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*. Clock drift is the phenomenon where the clock does not run at exactly the designated frequency. Normally, the effect is parameterized by the so-called *clock drift rate* that is given in *parts per million* (PPM) [16, 17]. For a drift rate r, the amount of drift at time t seconds is given by

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

where δ_t is in microseconds if r is in PPM.

Considering the models in Equation (4) and Equation (6), the direct result of clock drift is to create a mis-match between the received signal vector and the transmission matrix for each model. This is due to the fact that the received signal is generated from a different transmission matrix. This matrix is constructed from a transmission sequence where the pulse locations are shifted according to a certain drift rate. 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 in the error of $\boldsymbol{y}_d$.

Note that the ADC clock drift at the receiver can cause a similar effect to that caused by the transmitter 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 clock drift effect and it can be absorbed in the error matrix $\boldsymbol{\Delta}_A$.

In addition to the above two assumptions, it will also be assumed that the transmitter and receiver are

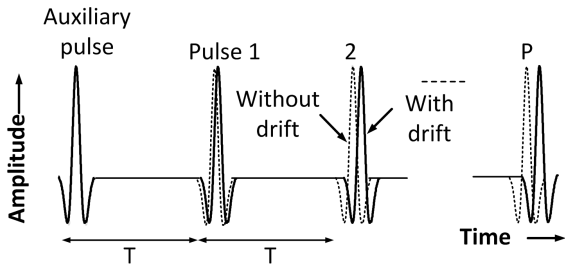

Fig. 4: An example of a transmit pulse train with clock drift (black). The original sequence without drift is depicted in grey colour.

perfectly synchronized at the time instant of the transmission of the first pulse in the pulse train. Clock drift effect starts from this point onward. Based on this assumption, the first pulse of the transmitted pulse train (the auxiliary pulse) will not suffer any time drift. The next pulse will be shifted forward or backward depending on whether the drift is positive or negative. Fig. 4 depicts an example for a transmitted pulse train with clock drift. The error matrix $\boldsymbol{\Delta}_A$ can be conceived as the difference between the two transmission matrices constructed from two different 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 known matrix $\boldsymbol{A}_d$.

The conclusion of this section is that, (16) is the practical model that will be considered in the sequel.

V. CHANNEL ESTIMATION

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

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

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 data contaminations. These estimators include Bayesian estimators such the *linear minimum mean squared error* (LMMSE) estimator, which, for zero-mean CIR and AWGN, is given by [18]

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

Note that the LMMSE approach assumes a stochastic CIR model contrary to the LS method that assumes a deterministic one. CIR models found in reality 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 (.ie., $\boldsymbol{C}_{hh}$ and σ_v^2) are not available in most practical cases[2]. 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:

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

To deal with the *model uncertainties* (also known as *error-in-variables*) in Equation (16), when no statistical information is available, a number of formalisms have been suggested. These include *total least squares* based methods [19, 20], H^∞ methods [21], and *bounded data uncertainty*(BDU) methods [22]. In this work, the BDU approach will be pursued to handle model uncertainties. The formalism is found to be well suited and can be applied in a computationally efficient manner, as will be demonstrated in this section.

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

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

where η is the upper bound on the 2-induced norm of $\boldsymbol{\Delta}_A$ and is assumed to be known, and η_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]

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

[2]CIR is usually changing in situations when UWB is used. Tracking these changes can be used for tracking the objects that cause them, which is an important UWB application.

Here $\boldsymbol{I}$ is the identity matrix of appropriate dimension, and γ_0 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 to the minimization problem in (21), η has to satisfy the inequality

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

Note that for $\gamma = 0$, the solution in (22) becomes 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 computation in an efficient manner.

V.I. Computing The Singular Values $\boldsymbol{\lambda}$

In a practical UWB system the matrix $\boldsymbol{A}_d$ is large and the calculation of its 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]. By exploiting the relationship between the singular values of the matrix $\boldsymbol{A}_d$ and the eigenvalues of the matrix $\boldsymbol{B} = \boldsymbol{A}_d^H \boldsymbol{A}_d$, we have

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

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

V.II. Finding γ

To find γ, (23) is iteratively solved. Since the right-hand-side of (23) is differentiable, the Newton's method [24] is one good option to carry out the task. Based on the Newton's method, the iterations for finding γ, 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 (27) reach convergence when $|\gamma[n+1] - \gamma[n]| < \epsilon$, where ϵ 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.

V.III. 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 using the FFT matrix $\boldsymbol{F}$ as

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

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

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

This equation shows that the process of CIR estimation under the BDU model can be achieved without incurring much computation. Note that while the matrix inverses needed for evaluating the LS, LMMSE and WLMMSE solutions can all be pre-computed, the BDU solution requires updating the matrix inverse $(\boldsymbol{A}_d^H \boldsymbol{A}_d + \gamma \boldsymbol{I})^{-1}$ according to the current received signal (γ 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 N.

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

1. Compute $\boldsymbol{\lambda}$, and then $\boldsymbol{\Sigma}$ based on (26).

2. Find γ using the iterations (27).

3. Compute $\hat{\boldsymbol{h}}_{BDU}$ 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 $\hat{\boldsymbol{h}}_{BDU}$ 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 η), which corresponds to solving the mini-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 Equation (22), the MSE is defined as

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

where $\boldsymbol{h}$ is the true CIR and $\hat{\boldsymbol{h}}$ is used instead of $\hat{\boldsymbol{h}}_{BDU}$ for simplification. 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:

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

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

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

where $\boldsymbol{\omega}$ is a vector that contains the eigenvalues of $\boldsymbol{A}_d^H \boldsymbol{\Delta}_A$, which is approximated as:

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

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:

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

where

$$\text{E}\left(\boldsymbol{y}_d \boldsymbol{y}_d^H\right) \dashrightarrow \sigma_v^2 \boldsymbol{I} + \sigma_h^2 \left(\boldsymbol{A}_d \boldsymbol{A}_d^H + \boldsymbol{A}_d \boldsymbol{\Delta}^H + \boldsymbol{\Delta}_A \boldsymbol{A}_d^H + \boldsymbol{\Delta}_A \boldsymbol{\Delta}_A^H\right). \tag{35}$$

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 it 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), MSE_{approx} equals $N\sigma_v^2$, which is the performance of the LS estimator [18].

(d) The benefit of the regularization parameter γ in reducing the noise effect is clear. It also reduces the contribution from the *imaginary* parameter of the drift. However, the way the regularization parameter is involved with the *real* parameter of the drift suggests that the regularization parameter may have a counter effect.

VI.I. Selecting η

First, recall that η 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 η that can be used. As the interest is customarily in the MSE performance, we can choose a value of η 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 η can then be obtained directly from (23). 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 η, we optimize the MSE performance under zero clock-drift conditions. By substituting $\boldsymbol{\Delta}_A = \mathbf{0}$ in (32), differentiating and setting the derivative equal to zero, we obtain

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

By substituting (36) in (23) solving for η, we obtain

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

From (36) and (37), it can clearly be seen that in the case where $\eta = \eta_{opt}$, the BDU estimator coincides with the WLMMSE estimator. In other words, by choosing η 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 ρ is very difficult to estimate accurately in most realistic scenarios. To circumvent this hurdle, we resort to approximate η_{opt} with its value at infinite SNR, i.e., when $\rho = 0$. This gives the final η that we propose to used in this paper:

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

Note that η_0 is determined only by the transmission matrix and can, therefore, be a priori calculated. 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 case.

2. It can be shown that η_0 increases with increase in both noise and drift. However, the increase in η_0 is *insignificant* since the MSE becomes flatter as noise or drift increases, as will be shown in the following section. Therefore, calculating η_0 under zero-noise zero-drift conditions can be considered as a sub-optimal choice.

VII. SIMULATIONS

In order to test the proposed equivalent-time sampling scheme 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 rate, $D = 5, 10, 20$ 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.

Fig. 5 (a) and (b) shows the variation of the MSE with η when a Gaussian pulse is used. Fig. 5 (a) is obtained from simulation, while Fig. 5 (b) is the approximate MSE (Equation (32). It can be seen that the analytical formulae provide a good approximation for the MSE. In spite of deviation in some case, the analytical formulae preserve the location of the optimal η with sufficient accuracy. The same phenomenon is seen in Fig. 5 (c) and (d), which are the counterparts of Fig. 5 (a) and (b) for a second derivative of a Gaussian pulse. The analytical formulae are giving a more accurate approximation of performance in the latter case. The effect of the SNR as a determining factor of performance is also more visible in the latter case than in the Gaussian pulse case. Note that the analytical performance appears to be somewhat optimistic in the case of low drift and low noise (e.g., for SNR=50 dB and $r = 1$ PPM). In each individual figure, the vertical dotted line marks the location of η_0 calculated using Equation (38). It can be seen that η_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 η_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. 5 that η_0 is different for different pulse shapes, which stresses the importance of (38) in choosing η. 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.

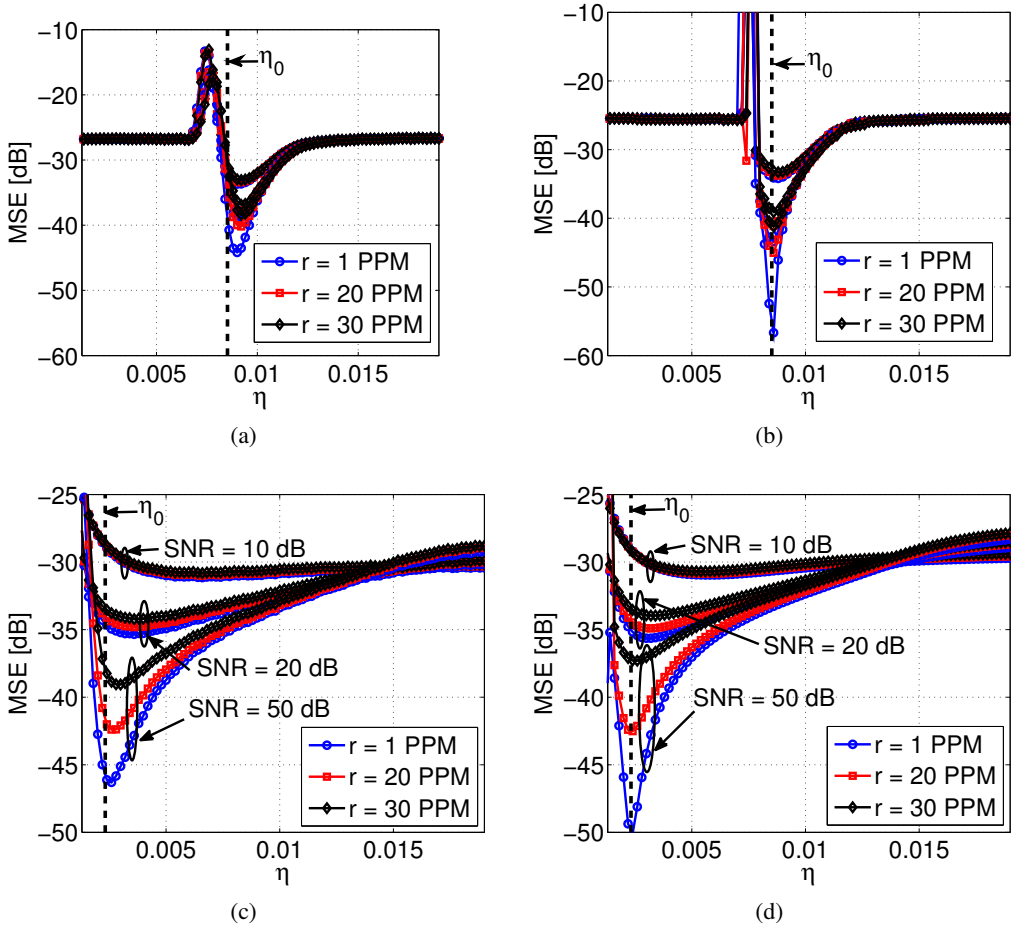

Fig. 5: MSE versus η: a) Gaussion pulse (simulation) b) Gaussian pulse (analytical) c) second derivative of Gaussian pulse (simulation) d) second derivative of Gaussian pulse (analytical)

Fig. 6 depicts the MSE versus the SNR for zero drift and sub-sampling factor $D = 1, 5, 10, 20$ and 50. The performance of the BDU, LS, WLMMSE and LMMSE estimators is shown. The $D = 1$ case represents the Nyquist sampling case where no sub-sampling is applied. It can be seen that for each individual estimator, the performance for each sub-sampling factor is exactly equivalent to that obtained at the Nyquist rate. The individual curves for the different D 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 in this case, that it assumes an error matrix $\mathbf{\Delta}_A \neq \mathbf{0}$, which is not correct. Subsequently, it will be shown that when clock drift is present, the BDU estimator can outperform the other three estimators.

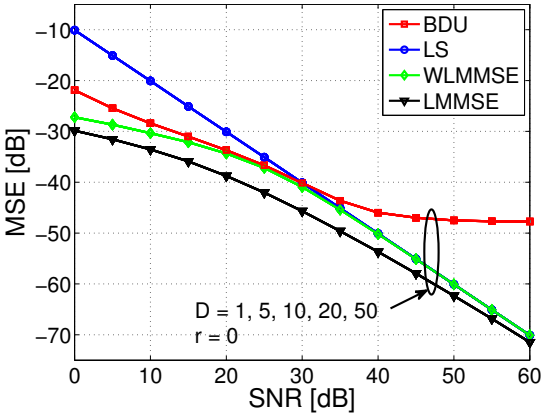

Fig. 6: MSE versus SNR for the BDU approach compared with the LS, WLMMSE and LMMSE approaches for zero clock drift and sub-sampling factor $D = 1, 5, 10, 20$ and 50 PPM.

Fig. 7 depicts the MSE versus the SNR for different sub-sampling factors, $D = 5, 10, 20$ and 50. The figure is plotted for a rather high drift rate of $r = 50$ PPM (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. Bothe 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 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 $D = 20$ and $D = 50$ 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 we are using more pulses. Consequently, the drift (the drift rate multiplied by the total duration, see Equation (15)) increases. This affect the other estimators more that the BDU, which is more equipped to cope with such situations of data uncertainty.

In Fig. 8, D is fixed to a value of 20, while r is varied to take the values 5, 10, 50 and 80 PPM.

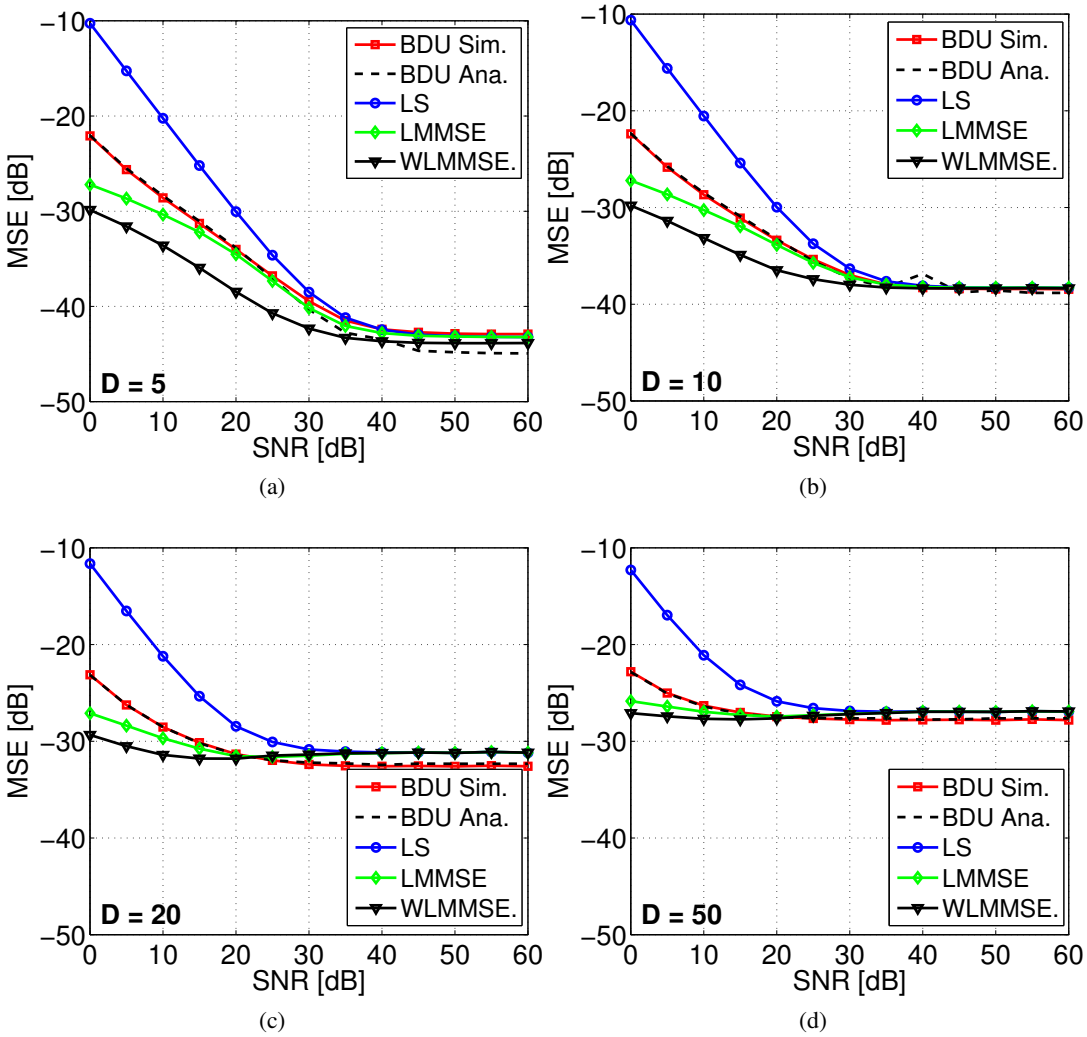

Fig. 7: MSE versus SNR for the BDU approach (simulation and analytical) compared with the LS, WLMMSE and LMMSE approaches for a drift rate $r = 50$ PPM and sub-sampling factor D equal to a) 5, b) 10, c) 20 and d) 50.

The low SNR performance of the four estimators, relative to one another, is similar to that in Fig. 7; performance is greatly determined by the SNR in this regime. Again, the BDU estimator performs better than the other estimators at high SNRs and low drift, as in the case of 50-PPM and 80-PPM drift rates.

From both Fig. 7 and Fig. 8, 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 9, which depicts performance against drift rate for $D = 20$ 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 applying a higher sub-sampling rate.

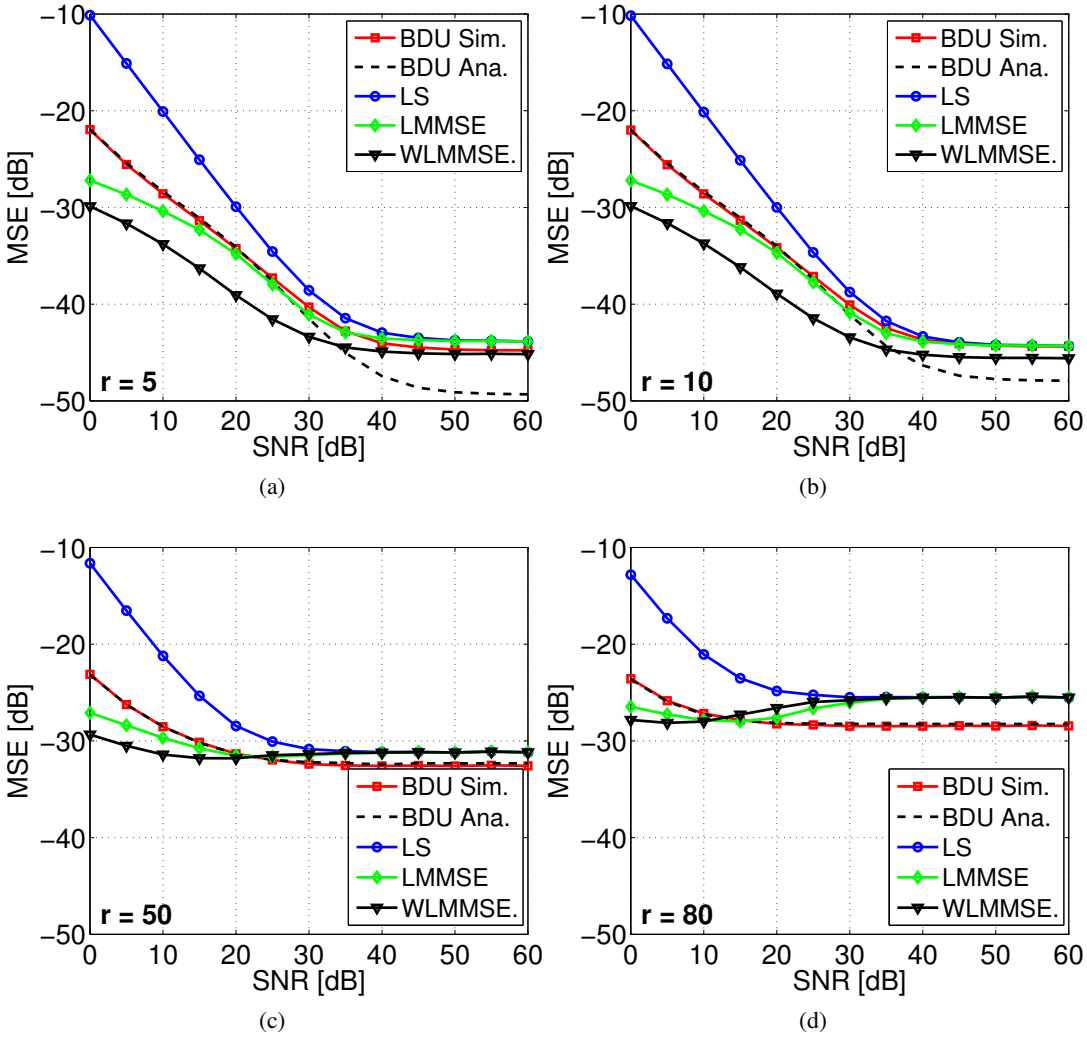

Fig. 8: MSE versus SNR for the BDU approach (simulation and analytical) compared with the LS, WLMMSE and LMMSE approaches for a sub-sampling factor $D = 20$ and the drift rate r equal to a) 5, b) 10, c) 50 and d) 80 PPM.

VIII. CONCLUSIONS

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

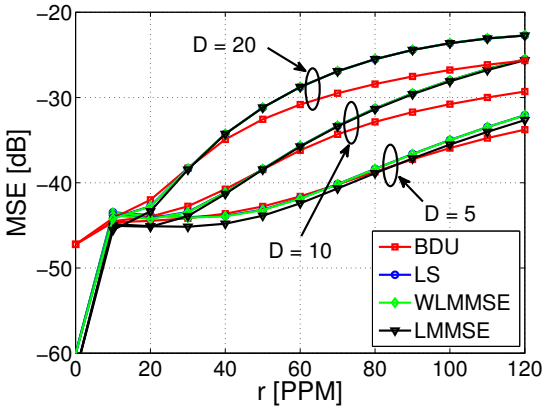

Fig. 9: MSE versus drift rate r for the BDU approach compared with the LS, WLMMSE and LMMSE approaches for SNR = 50 dB and sub-sampling factor $D = 5, 10$ and 20.

observations, and that the length of 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.

Appendices

A. DERIVATION OF THE MSE

Starting from Equation (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

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

Using (16), the expectation can be expanded to

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

where

$$\boldsymbol{C}_{hh} = \mathrm{E}\left(\boldsymbol{h}\boldsymbol{h}^H\right) \qquad \text{(A.3)}$$

is the CIR covariance matrix.

The second term of the MSE is

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

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

$$\begin{aligned}\mathrm{tr}\left[\mathrm{E}\left(\boldsymbol{h}\hat{\boldsymbol{h}}^H\right)\right] &= \mathrm{tr}\left[\mathrm{E}\left(\boldsymbol{h}\boldsymbol{y}_d^H\right)\boldsymbol{A}_d\left(\boldsymbol{A}_d^H\boldsymbol{A}_d + \bar{\gamma}\boldsymbol{I}\right)^{-1}\right] \\ &= \mathrm{tr}\left[\mathrm{E}\left(\boldsymbol{h}\boldsymbol{h}^H\boldsymbol{A}_d^H + \boldsymbol{h}\boldsymbol{h}^H\boldsymbol{\Delta}_A^H + \boldsymbol{h}\boldsymbol{v}_d^H\right)\boldsymbol{A}_d\left(\boldsymbol{A}_d^H\boldsymbol{A}_d + \bar{\gamma}\boldsymbol{I}\right)^{-1}\right] \\ &= \mathrm{tr}\left[\left(\boldsymbol{C}_{hh}\boldsymbol{A}_d^H\boldsymbol{A}_d + \boldsymbol{C}_{hh}\boldsymbol{\Delta}_A^H\boldsymbol{A}_d\right)\left(\boldsymbol{A}_d^H\boldsymbol{A}_d + \bar{\gamma}\boldsymbol{I}\right)^{-1}\right]. \qquad \text{(A.5)}\end{aligned}$$

Now, substituting (A.1)–(A.5) in (30), we obtain the expression for the *exact* MSE as

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

To obtain a simpler expression, we replace the CIR covariance by its *white* equivalent

$$\boldsymbol{C}_{hh} = \sigma_h^2\boldsymbol{I}. \qquad \text{(A.7)}$$

This, after manipulation, results in the white MSE

$$\begin{aligned}\text{MSE}_{white} &= \sigma_v^2 \text{tr}\left[\left(\boldsymbol{A}_d^H\boldsymbol{A}_d+\bar{\gamma}\boldsymbol{I}\right)^{-1}\boldsymbol{A}_d^H\boldsymbol{A}\left(\boldsymbol{A}_d^H\boldsymbol{A}_d+\bar{\gamma}\boldsymbol{I}\right)^{-1}\right] \\ &+ \sigma_h^2 \text{tr}\Big[\left(\boldsymbol{A}_d^H\boldsymbol{A}_d+\bar{\gamma}\boldsymbol{I}\right)^{-1} \\ &\qquad \left(\boldsymbol{A}^H\boldsymbol{\Delta}\boldsymbol{\Delta}^H\boldsymbol{A}-\bar{\gamma}\boldsymbol{A}^H\boldsymbol{\Delta}-\bar{\gamma}\boldsymbol{\Delta}^H\boldsymbol{A}+\bar{\gamma}^2\boldsymbol{I}\right)\left(\boldsymbol{A}_d^H\boldsymbol{A}_d+\bar{\gamma}\boldsymbol{I}\right)^{-1}\Big]. \quad \text{(A.8)}\end{aligned}$$

Further to the previous whiteness approximation, now, we introduce the following replacements: Now, let us introduce the following approximations:

$$\begin{aligned}\boldsymbol{A}_d^H\boldsymbol{\Delta}_A &\dashrightarrow \boldsymbol{F}\boldsymbol{\Omega}\boldsymbol{F}^H \\ \boldsymbol{\Delta}_A^H\boldsymbol{A}_d &\dashrightarrow \boldsymbol{F}\boldsymbol{\Omega}^H\boldsymbol{F}^H, \quad \text{(A.9)}\end{aligned}$$

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

$$\boldsymbol{\Omega} = \text{diag}\left[\text{diag}\left(\boldsymbol{F}^H\boldsymbol{A}_d^H\boldsymbol{\Delta}_A\boldsymbol{F}\right)\right]. \quad \text{(A.10)}$$

Note that (A.10) 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 $\tilde{\boldsymbol{A}}_d$ and the matrix $\boldsymbol{A}_d$. The matrix $\tilde{\boldsymbol{A}}_d$ is obtained by decimating a matrix $\tilde{\boldsymbol{A}}$, 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 $\boldsymbol{A}_d^H\boldsymbol{\Delta}_A = \boldsymbol{A}_d^H\tilde{\boldsymbol{A}}_d - \boldsymbol{A}_d^H\boldsymbol{A}_d$ is approximately similar to that of $\boldsymbol{A}_d^H\boldsymbol{A}_d$, including the circulant property, which justifies (A.9) and (A.10).

Now, substituting (A.9) together with (28) in (A.8) and manipulating based on trace and FFT matrix properties results in the *approximate* MSE

$$\begin{aligned}\text{MSE}_{approx} &= \sigma_v^2 \text{tr}\Big[\boldsymbol{\Sigma}^2\left(\boldsymbol{\Sigma}^2+\bar{\gamma}\boldsymbol{I}\right)^{-2}\Big] \\ &+ \sigma_h^2 \text{tr}\Big\{\left(\boldsymbol{\Sigma}^2+\bar{\gamma}\boldsymbol{I}\right)^{-2}\left[\bar{\gamma}^2\boldsymbol{I}-2\bar{\gamma}\Re(\boldsymbol{\Omega})+|\boldsymbol{\Omega}|^2\right]\Big\}. \quad \text{(A.11)}\end{aligned}$$

Manipulating further yields (32).

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

$$\bar{\gamma}[n+1] = \bar{\gamma}[n] + \frac{\text{tr}\left[\boldsymbol{U}(\boldsymbol{\Sigma}^2-\eta^2\boldsymbol{I})(\boldsymbol{\Sigma}^2+\bar{\gamma}[n]\boldsymbol{I})^{-2}\boldsymbol{U}^H\text{E}(\boldsymbol{y}_d\boldsymbol{y}_d^H)\right]}{2\text{tr}\left[\boldsymbol{U}(\boldsymbol{\Sigma}^2-\eta^2\boldsymbol{I})(\boldsymbol{\Sigma}^2+\bar{\gamma}[n]\boldsymbol{I})^{-3}\boldsymbol{U}^H\text{E}(\boldsymbol{y}_d\boldsymbol{y}_d^H)\right]}, \quad \text{(A.12)}$$

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

$$\mathrm{E}(\boldsymbol{y}_d\boldsymbol{y}_d^H) = \sigma_v^2\boldsymbol{I} + \sigma_h^2\left(\boldsymbol{A}_d\boldsymbol{A}_d^H + \boldsymbol{A}_d\boldsymbol{\Delta}^H + \boldsymbol{\Delta}_A\boldsymbol{A}_d^H + \boldsymbol{\Delta}_A\boldsymbol{\Delta}_A^H\right). \tag{A.13}$$

B. REFERENCES

[1] G. B. Giannakis L. Yang, "Ultra-wideband communications: An idea whose time has come," *IEEE Signal Processing Magazine*, 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 Journal of Selected Topics in Signal Processing*, 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 Transactions on Wireless Communications*, 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 Journal on Selected Areas in Communications*, vol. 20, no. 9, pp. 1638–1645, Dec 2002.

[5] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information 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 Personal Area Networks, P802.15-02/490r1-SG3a*, Feb 2003.

[7] I. Maravić, J. Kusuma, and M. Vetterli, "Low-sampling rate uwb channeln characterization and synchronization," *Journal of Communications and Networks*, vol. 5, no. 2, pp. 319327, 2003.

[8] Zhengdao Wang and Xiaofan Yang, "Ultra wide-band communications with blind channel estimation based on first-order statistics," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP04)*, May 2004, vol. 4, pp. 529–532.

[9] Ye (Geoffrey) Li, Andreas F. Molisch, and Jinyun Zhang, "Channel estimation and signal detection for uwb," in *Proc. IEEE WPMC*, 2003.

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

[11] K. Lawton and E. McConnell, *Equivalent Time Sampling for High-Speed Repetitive Signals Using E Series Boards and NI-DAQ Software, Application Note 066*, National Instruments Corporation, 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 *16th International Conference on Electronics, Communications and Computers (CONIELECOMP 2006)*, 2006, pp. 34–34.

[13] S. Gezici, "Theoretical limits for estimation of periodic movements in pulse-based uwb systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 3, pp. 405–417, 2007.

[14] Yuan Zhou, Yong Liang Guan, and C.L. Law, "Modified phase-only correlator with kurtosis-based amplified-noise suppression," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3341–3345, 2010.

[15] Andreas F. Molisch, Kannan Balakrishnan, Chia chin Chong, Shahriar Emami, Andrew Fort, Johan Karedal, Juergen Kunisch, Hans Schantz, Ulrich Schuster, and Kai Siwiak, "Ieee 802.15.4a channel model - final report," in *Converging: Technology, work and learning. Australian Government Printing Service*. 2004, Available Online.

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

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

[18] S. M. Kay, *Fundamentals of Statistical Signal Processing*, Printice Hall, 1993.

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

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

[21] B. Hassibi, A. Sayed, and T. Kailath, *Indeffinite Quadratic Estimation and Control: A Uniffied Approach to H^2 and H^∞ Theories*, SIAM, Philadelphia, 1999.

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

[23] M. Chen, "On the solution of circulant linear systems," *SIAM Journal on Numerical Analysis*, vol. 24, no. 3, pp. 668–683, Jun. 1987.

[24] C. Zarowski, *An Introduction to Numerical Analysis for Electrical and Computer Engineers*, John Wiley & Sons, 2004.

[25] M. Benedetto, T. Kaiser, A. Molisch, I. Oppermann, C. Politano, and D. Porcino, *UWB Communication Systems–A Comprehensive Overview*, Hindawi, 2006.