

Improved Maximum Likelihood Frequency Offset Estimation Based on Likelihood Metric Design

Hlaing Minn*, *Member, IEEE* and Poramate Tarasak, *Member, IEEE*

Abstract—For emerging high data-rate communication systems in highly dispersive channels such as ultra-wideband systems, possible frequency offsets could be larger than the estimation range of the existing methods using training signals with identical parts or repetitive training signals (i.e., the training signals are composed of several identical sub-blocks or are obtained by repeating a training sub-block for several times). This paper presents a novel improved maximum likelihood frequency offset estimator which can handle at least twice the estimation range of the existing methods using training signals with identical parts and achieves a better estimation performance. Based on the likelihood metric, a new design metric is introduced which is a pair-wise error probability (PEP) between the correct frequency offset point and a trial frequency offset point. The proposed PEP metric gives more theoretical insights on the performance of practical maximum likelihood estimators. How to design the PEP to achieve both a larger estimation range and a better estimation performance in fading channel environments is also presented and the corresponding estimator implementation is described.

Keywords—synchronization, estimation, frequency offset, likelihood metric design.

I. INTRODUCTION

Frequency synchronization is an essential task at a communication receiver. For packet-based systems such as 2G, 3G and beyond-3G cellular systems, wireless LANs, wireless MANs, etc., training signal based frequency offset estimation and compensation are typically performed. The scope of this paper encompasses all packet-based single-carrier as well as multi-carrier systems except CDMA-based systems. There are several existing works on frequency offset estimation, e.g., [1]- [11]. They are mainly based on a correlation term of the training signal in time-domain or frequency-domain [1]- [4], maximum-likelihood principle [5] [6], a Bayesian approach [7], a combination of correlation terms in a sub-optimal way [8], or a combination of correlation terms using best linear unbiased estimation principle [9]- [11]. Most of them employ repetitive training signals consisting of several identical parts (or in the form of cyclic prefixes) which also yield low complexity estimators.

In [9], a maximum likelihood frequency offset estimation method (MLE#1) was presented based on a joint estimation of frequency offset and channel impulse response. With a proper training signal, MLE#1 can handle absolute frequency offsets

less than half of the symbol rate which is the maximum possible estimation range for any estimator operating on symbol-rate received signal samples. To reduce the MLE#1's very high complexity, [9] also presented MLE#2 which utilized a periodic training signal with a period of L samples (the number of channel taps). The complexity of MLE#2 is approximately $1/L$ times that of MLE#1 but the corresponding estimation range of MLE#2 is reduced to $1/(2L)$ times the symbol rate. Since the estimation range of MLE#2 is inversely proportional to the number of channel taps, MLE#2 cannot be applied to systems where possible frequency offsets are larger than the above range. For example, in ultra-wideband (UWB) systems, the number of channel taps can be quite large and possible frequency offsets (due to very high carrier frequency and/or low cost devices) can be larger than the estimation range of MLE#2. Similarly, the above-mentioned existing methods with repetitive training signals experience the same problem.

For highly-dispersive channel environments, developing frequency offset estimators which can handle all possible frequency offsets with reasonable complexity is a challenging problem which has not been addressed in the literature. As UWB systems become more prominent, the above problem becomes an important issue. Hence, in this paper, we address this issue and develop a novel improved maximum likelihood frequency offset estimation method which can handle a larger frequency offset with a comparable complexity. Based on the likelihood metric, a new design metric named pair-wise error probability (PEP) is introduced. The training signal consisting of several identical sub-blocks are designed based on the PEP metric in order to achieve the estimation range extension. The proposed design also brings in estimation performance improvement. The proposed scheme can be applied to single-carrier as well as multi-carrier systems except CDMA systems¹.

The proposed approach can be related to the optimal periodic training signal design for frequency offset estimation presented in [12]. The approach from [12] is based on the Cramer-Rao lower bound (CRB) and it does not provide information on whether a practical frequency offset estimator will achieve the improvement projected in the CRB. On the other hand, the proposed approach in this paper is based on the likelihood metric of a practical maximum likelihood estimator. It provides more insights for practical maximum likelihood estimators and it ensures the improvement in practice.

The rest of the paper is organized as follows. Section II describes signal model and Section III presents the proposed

This work was supported in part by the Erik Jonsson School Research Excellence Initiative, the University of Texas at Dallas. H. Minn is with the Department of Electrical Engineering, School of Engineering and Computer Science, University of Texas at Dallas, Richardson, TX 75083-0688, U.S.A. (hlaing.minn@utdallas.edu). P. Tarasak is with the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, Korea, (ptarasak@stein.kaist.ac.kr).

*Contact Author

¹For CDMA systems, the proposed concept could still be applied in down-link but multi-user interference and multi-user training signal designs would also need to be taken into consideration in uplink.

approach for improved maximum likelihood estimator with extended estimation range. Simulation results and discussions are given in Section IV and finally the paper is concluded in Section V.

II. SIGNAL MODEL

For complexity reduction, consider an arbitrary training signal consisting of several identical sub-blocks (say, V sub-blocks). Each training sub-block is composed of L training samples² $\{s_k : k = 0, \dots, L - 1\}$ where L is the number of sample-spaced channel taps. The locations of the identical sub-blocks are defined by the time-indices of the first samples of the sub-blocks, namely m_0, m_1, \dots, m_{V-1} where $m_n - m_{n-1} \geq L$ for $n = 1, \dots, V - 1$. If two adjacent sub-blocks are not consecutively located, i.e., $m_n - m_{n-1} > L$, then there can be null samples or non-zero samples between the two sub-blocks. The non-zero samples between the two training sub-blocks could convey some control information or could be used as training samples for other synchronization tasks. The first sub-block serves as a cyclic prefix (CP). Similarly, the n -th sub-block will serve as a CP if $m_n - m_{n-1} > L$. At the receiver, the observation vector for frequency offset estimation is formed from the received training samples by removing the CPs (the first sub-block and all other training sub-blocks $\{n\}$ for which $m_n - m_{n-1} > L$) and the null or data samples between any two training sub-blocks. Suppose in the observation vector there are U sub-blocks with the corresponding time-indices $\{l_0, l_1, \dots, l_{U-1}\}$ where $l_i \in \{m_n\}$. Fig. 1 depicts several training signal structures and corresponding construction of observation vectors (composed of $\{l_k\}$) where shaded or unshaded blocks represent transmitted training signal sub-blocks and blank spaces between training sub-blocks represent null or non-zero samples for other purposes.

Consider a complex baseband received observation vector \mathbf{r} (composed of the received training sub-blocks with time-indices $\{l_k\}$) given by

$$\mathbf{r} = [r(0), r(1), \dots, r(UL - 1)]^T \quad (1)$$

where the indices of $\{r(k)\}$ are with reference to the observation vector (not the received sample time-indices). Then we can express \mathbf{r} as

$$\mathbf{r} = \Gamma(v)\mathbf{S}\mathbf{h} + \mathbf{w} = \Gamma(v)\mathbf{x} + \mathbf{w} \quad (2)$$

where

$$\mathbf{h} = [h(0), h(1), \dots, h(L - 1)]^T \quad (3)$$

$$\mathbf{w} = [w(0), w(1), \dots, w(UL - 1)]^T \quad (4)$$

$$\Gamma(v) = \text{diag}\{e^{j\theta_0(v)}\mathbf{W}(v), e^{j\theta_1(v)}\mathbf{W}(v), \dots, e^{j\theta_{U-1}(v)}\mathbf{W}(v)\} \quad (5)$$

$$\mathbf{W}(v) = [1, e^{j2\pi v}, \dots, e^{j2\pi(L-1)v}] \quad (6)$$

$$\theta_k(v) = 2\pi l_k v \quad (7)$$

$$[\mathbf{S}]_{k,n} = s_{(k-n)_L}, 0 \leq k \leq UL - 1, 0 \leq n \leq L - 1. \quad (8)$$

²We will use sample and symbol interchangeably.

In the above equations, v is the carrier frequency offset normalized by the sample rate $1/T_s$, $\{w(n)\}$ are independent and identically distributed (iid) zero-mean circularly-symmetric complex Gaussian noise samples each having a variance of σ_n^2 , $\{h(k)\}$ are the channel tap gains assumed to remain constant during the training block, and $(\cdot)_L$ denotes a modulo- L operation. The superscripts $*$, T , and H represent the conjugate, the transpose and the conjugate transpose operations, respectively.

Note that \mathbf{S} consists of U identical $L \times L$ sub-matrices which are designed to be of full rank. In practice, the exact number of channel taps may not be known and it can vary as well. Hence, L should represent an upper bound of the number of channel taps. In this case, the channel vector \mathbf{h} contains actual channel taps appended with zero-value taps and the signal model remains the same.

III. THE PROPOSED METHOD

We consider a joint estimation of \mathbf{h} and v as in [9]. The likelihood function is given by

$$\Lambda(\mathbf{r}; \tilde{\mathbf{h}}, \tilde{v}) = \frac{1}{(\pi\sigma_n^2)^{UL}} \exp\left\{-\frac{1}{\sigma_n^2} \|\mathbf{r} - \Gamma(\tilde{v})\mathbf{S}\tilde{\mathbf{h}}\|^2\right\} \quad (9)$$

where $\tilde{\mathbf{h}}$ and \tilde{v} are trial values of \mathbf{h} and v in finding/searching the best values (of $\tilde{\mathbf{h}}$ and \tilde{v}) that maximize the likelihood function and $\|\cdot\|^2$ represents a Euclidean norm-square. For a fixed \tilde{v} , we find $\tilde{\mathbf{h}}$ that maximizes the likelihood function $\Lambda(\mathbf{r}; \tilde{\mathbf{h}}, \tilde{v})$ by differentiating $\|\mathbf{r} - \Gamma(\tilde{v})\mathbf{S}\tilde{\mathbf{h}}\|^2$ with respect to $\tilde{\mathbf{h}}$ and equating the result to zero. The corresponding estimate $\hat{\mathbf{h}}(\tilde{v})$ is given by

$$\hat{\mathbf{h}}(\tilde{v}) = (\mathbf{S}^H\mathbf{S})^{-1}\mathbf{S}^H\Gamma^H(\tilde{v})\mathbf{r} \quad (10)$$

where \mathbf{S} has been designed to have a full rank. After substituting (10) into (9), the maximum likelihood estimate of v is given by

$$\hat{v} = \arg \max_{\tilde{v}} \{g(\tilde{v})\} \quad (11)$$

where

$$g(\tilde{v}) = \mathbf{r}^H \Gamma(\tilde{v})\mathbf{B}\Gamma^H(\tilde{v})\mathbf{r}. \quad (12)$$

In the above equation, \mathbf{B} is a $UL \times UL$ projection matrix given by

$$\mathbf{B} = \mathbf{S}(\mathbf{S}^H\mathbf{S})^{-1}\mathbf{S}^H = \frac{1}{U} \begin{bmatrix} \mathbf{I}_L & \dots & \mathbf{I}_L \\ \vdots & \ddots & \vdots \\ \mathbf{I}_L & \dots & \mathbf{I}_L \end{bmatrix} \quad (13)$$

where \mathbf{I}_L is an $L \times L$ identity matrix. The above derivation ((9)-(13)) is the same as [9] except that $\Gamma(\tilde{v})$ and the formation of the observation vector \mathbf{r} are different from [9]. After some simplification and dropping unnecessary factors, we obtain

$$g(\tilde{v}) = \Re \left\{ \sum_{m=0}^{U-1} \sum_{l=m}^{U-1} \rho(m, l) e^{j(\theta_m(\tilde{v}) - \theta_l(\tilde{v}))} \right\} = \Re \{ \Omega(\tilde{v})\boldsymbol{\rho} \} \quad (14)$$

where $\Re\{X\}$ denotes the real part of X and

$$\rho(m, l) = \sum_{n=0}^{L-1} r(lL + n) r^*(mL + n) \quad (15)$$

$$\rho = [\rho(0, 0), \rho(0, 1), \dots, \rho(0, U-1), \rho(1, 1), \dots, \rho(1, U-1), \rho(2, 2), \dots, \rho(U-1, U-1)]^T \quad (16)$$

$$\Omega(\tilde{v}) = [e^{j(\theta_0(\tilde{v}) - \theta_0(\tilde{v}))}, e^{j(\theta_0(\tilde{v}) - \theta_1(\tilde{v}))}, \dots, e^{j(\theta_0(\tilde{v}) - \theta_{U-1}(\tilde{v}))}, e^{j(\theta_1(\tilde{v}) - \theta_1(\tilde{v}))}, \dots, e^{j(\theta_1(\tilde{v}) - \theta_{U-1}(\tilde{v}))}, e^{j(\theta_2(\tilde{v}) - \theta_2(\tilde{v}))}, \dots, e^{j(\theta_{U-1}(\tilde{v}) - \theta_{U-1}(\tilde{v}))}] \quad (17)$$

Now, we introduce a new variable to be used in developing our proposed method as

$$Z(\tilde{v}) = g(v) - g(\tilde{v}) = \Re\{[\Omega(v) - \Omega(\tilde{v})]\rho\} \quad (18)$$

We will design the likelihood metric to achieve a larger estimation range and a better estimation performance. This will be accomplished by means of the design on the training sub-block locations. We will use $P[Z(\tilde{v}) \leq 0]$ as our performance measure in the design. This measure indicates how likely a trial point \tilde{v} will be chosen as the frequency offset estimate between the actual frequency offset v and the trial point \tilde{v} . In other words, this probability can be considered as a pairwise error probability (PEP) of frequency offset estimation where the exact frequency offset and any *other* frequency offset trial point constitute the pair. Note that when $\tilde{v} = v$, this probability equals to one and it does not represent PEP but for convenience, $P[Z(\tilde{v}) \leq 0]$ will be referred to as the PEP metric throughout the paper.

We can approximate $Z(\tilde{v})$ as a Gaussian random variable (see Appendix-A for justification). The mean and the variance of $Z(\tilde{v})$ are, respectively, given by

$$E[Z(\tilde{v})] = \frac{1}{2} [(\Omega(v) - \Omega(\tilde{v}))E[\rho] + E[\rho^H](\Omega(v) - \Omega(\tilde{v}))^H] \quad (19)$$

$$\begin{aligned} \text{Var}[Z(\tilde{v})] = & \frac{1}{2} (\Omega(v) - \Omega(\tilde{v}))E[(\rho - E[\rho])(\rho - E[\rho])^H](\Omega(v) - \Omega(\tilde{v}))^H \\ & + \frac{1}{4} (\Omega(v) - \Omega(\tilde{v}))E[(\rho - E[\rho])(\rho - E[\rho])^T](\Omega(v) - \Omega(\tilde{v}))^T \\ & + \frac{1}{4} (\Omega(v) - \Omega(\tilde{v}))^*E[(\rho - E[\rho])^*(\rho - E[\rho])^H](\Omega(v) - \Omega(\tilde{v}))^H \end{aligned} \quad (20)$$

Then our design performance measure, PEP, is given by

$$P[Z(\tilde{v}) \leq 0] = Q\left(\frac{E[Z(\tilde{v})]}{\text{Std}[Z(\tilde{v})]}\right) \quad (21)$$

where $\text{Std}[X]$ denotes the standard deviation of X and $Q(\cdot)$ is the Gaussian tail probability. One can note from the above equation that $E[Z(\tilde{v})]/\text{Std}[Z(\tilde{v})]$ can be used as a design performance measure as well.

Next, we calculate the mean and the correlation of $\{\rho(k, l)\}$ required in (19) and (20). After a direct (but lengthy) calculation, we obtain

$$E[\rho(k, l)] = E_1 + L\sigma_n^2\delta[k - l] \quad (22)$$

$$\begin{aligned} E[\rho^*(k, l)\rho(m, n)] = & E_1^2 + LE_1\sigma_n^2(\delta[m - n] + \delta[l - k]) \\ & + E_1\sigma_n^2(\delta[k - m] + \delta[l - n]) + f(\mathbf{w}, l, k, m, n) \end{aligned} \quad (23)$$

$$\begin{aligned} E[\rho(k, l)\rho(m, n)] = & E_1^2 + LE_1\sigma_n^2(\delta[m - n] + \delta[l - k]) \\ & + E_1\sigma_n^2(\delta[l - m] + \delta[k - n]) + f(\mathbf{w}, k, l, m, n) \end{aligned} \quad (24)$$

where $E_1 = \sum_{k=0}^{L-1} |x_k|^2$ and

$$f(\mathbf{w}, l, k, m, n) = \begin{cases} (2L + 2\{(L-1)!\})\sigma_n^4, & \text{if } l = k = m = n \\ L^2\sigma_n^4, & \text{if } l = k, m = n, l \neq m \\ L\sigma_n^4, & \text{if } k = m, l = n, k \neq l \\ 0, & \text{otherwise.} \end{cases} \quad (25)$$

Note that in developing our design measure we have considered a fixed channel output sub-block energy E_1 or a fixed instantaneous (snap-shot) signal-to-noise ratio $\text{SNR}_i = E_1/(L\sigma_n^2)$. In practical fading channels, E_1 or SNR_i will fluctuate. For the estimation to be robust in fading environments, we will use $\text{SNR}_i = 0$ dB in our design. Different values of actual normalized frequency offset v simply result in shifted versions (in the \tilde{v} -axis) of the metrics $Z(\tilde{v})$, $E[Z(\tilde{v})]/\text{Std}[Z(\tilde{v})]$, and $P[Z(\tilde{v}) \leq 0]$ but they do not change the shapes of the metrics (see Appendix-A for the proof). The shifting of the metrics will not affect the frequency offset estimation range which is determined by the distance in \tilde{v} -axis between the metric peak corresponding to the actual frequency offset and the adjacent metric peak with comparable metric value (metric nulls in place of metric peaks for the metric $E[Z(\tilde{v})]/\text{Std}[Z(\tilde{v})]$). Hence, in our likelihood metric design, without loss of generality, we set $v = 0$. By plotting $P[Z(\tilde{v}) \leq 0]$ versus \tilde{v} for different training sub-block locations, one can design the training sub-block locations that achieve a larger frequency offset estimation range (without ambiguity) and a better frequency offset estimation mean-square error (MSE) performance. The expected value of $g(\tilde{v})$ or the ratio $E[Z(\tilde{v})]/\text{Std}[Z(\tilde{v})]$ can also be used as the design performance measures. Using $P[Z(\tilde{v}) \leq 0]$ is more informative in that it indicates how likely a trial frequency offset point \tilde{v} will be chosen as opposed to the correct frequency offset v .

To illustrate our proposed approach, let us consider a system where $V = 10$ identical training sub-blocks are transmitted and $L = 8$ (see Fig. 1). The observation vector will have a smaller number of sub-blocks U (less training energy used in the estimation) if there are more non-consecutive groups of sub-blocks, i.e., if there are more k for which $m_k - m_{k-1} > L$. Hence, we consider a scheme with two non-consecutive groups of training sub-blocks (there is only one k for which $m_k - m_{k-1} > L$). The two groups are separated by d samples. The first group contains $K_1 + 1$ sub-blocks and the second has $U - K_1 + 1$ sub-blocks. The 0-th and $(K_1 + 1)$ -th sub-blocks serve as CPs and are removed at the receiver. Hence, if $d > 0$, the observation vector has $U = 8$ training sub-blocks with the corresponding time indices³ given by $l_i = iL$ for $i = 0, 1, \dots, K_1 - 1$, and $l_{K_1+n} = d + (K_1 + 1 + n)L$ for $n = 0, \dots, U - K_1 - 1$. If $d = 0$, the observation vector contains $U = 9$ consecutive sub-blocks with the corresponding time indices $l_i = iL$ for $i = 0, 1, \dots, U - 1$, which is the conventional training structure used in [9]- [11]. Our objective in this illustrative example is to find d which gives a larger frequency offset estimation range and a better estimation MSE than the conventional approaches (corresponding to $d = 0$).

³The time index of the first transmitted sub-block is assumed to be $-L$.

In Figs. 2-9, we present effects of d , K_1 , U , and SNR_i on several metrics. In Fig. 2, the normalized mean values of the likelihood metric $g(\tilde{v})$ for d from 0 to $L - 1$ are presented where $U = 9$ for $d = 0$ and $U = 8$, $K_1 = 4$ for $d > 0$. The corresponding plots of $E[Z(\tilde{v})]/\text{Std}[Z(\tilde{v})]$ and PEP are shown in Figs. 3 and 4, respectively. The metric lobe centered around $\tilde{v} = 0$ (correct frequency offset) will be called mainlobe and the other lobes will be referred to as sidelobes. For the metric $E[Z(\tilde{v})]/\text{Std}[Z(\tilde{v})]$, the metric null at $\tilde{v} = 0$ (correct frequency offset) will be called main-null and the other nulls will be referred to as side-nulls. At $\tilde{v} = k/L$ where k is an integer, the mean values of the likelihood metric are the same for $d = 0$ hence limiting the estimation range of the conventional approaches to $|v| < 1/(2L)$. Similarly, the values of $E[Z(\tilde{v})]/\text{Std}[Z(\tilde{v})]$ are all zeros at $\tilde{v} = k/L$ (is 0/0 at $\tilde{v} = v = 0$) and the PEPs are all 0.5 at $\tilde{v} = k/L$ (is 1 at $\tilde{v} = v = 0$). In other words, for $d = 0$, the metric sidelobe peaks (or side-nulls) have the same (or almost the same) values as the mainlobe peak (or main-null) and this fact limits the unambiguous frequency offset estimation range.

For $d > 0$ in Figs. 2-4, the values of all metrics (the mean of likelihood metric, $E[Z(\tilde{v})]/\text{Std}[Z(\tilde{v})]$, and the PEP) change at some or all points of $\tilde{v} = k/L$, ($k \neq 0$), hence opening up the possibility of estimation range extension. In other words, the sidelobe peaks (or side-nulls) adjacent to the mainlobe peak (or main-null) take on values which are sufficiently distanced from the mainlobe-peak value (or side-null value), hence increasing the unambiguous frequency offset estimation range. For example, for $d = 2$ or 6, the estimation range becomes 4 times that of $d = 0$ but due to relatively large sidelobe peaks within the range, its estimation performance could be affected at low SNR. Of particular interest is $d = L/2$ case whose estimation range is twice of the range with $d = 0$ and its sidelobe peaks within the range are relatively small, hence ensuring high accuracy of estimation.

The effects of $d = kL/2$ for some integer k are presented in Figs. 5-7. Note that $d = 0$ corresponds to the conventional structure [9] [11] [13] and $d = mL$ for a nonzero positive integer m corresponds to the structures considered in [12]. For these structures (corresponding to $d = 0, 16, 32$ in the figures), the estimation range is $|v| < 1/(2L)$ and a larger d gives a sharper metric mainlobe (or main-null) resulting in a better estimation MSE performance⁴. However, a larger d introduces new sidelobes of the likelihood or PEP metric whose values increase as d increases (new side-nulls for the $E[Z(\tilde{v})]/\text{Std}[Z(\tilde{v})]$ metric whose values decrease as d increases), which may limit the use of a very large d for low SNR.

We observe that the training structure design by minimizing the CRB as in [12] is in fact making the likelihood metric sharper around the correct frequency offset. On the other hand, the training design by minimizing the CRB may not reveal the feasibility of the maximum likelihood estimator. For example, with a very large d at low SNR, the maximum likelihood estimator would not give a reliable result due to large sidelobe

peaks of the likelihood metric within the estimation range. This fact cannot be deduced from the minimum CRB design from [12]. For $d = kL/2$ with an odd integer k (i.e., $d = 4, 20$, and 36 in Figs. 5-7), the estimation range is doubled and similar discussion applies – a larger d gives a sharper likelihood metric mainlobe but a very large d may not be give a reliable estimate due to the increased sidelobe peaks.

Next, the impacts of different K_1 values on the PEP are shown in Fig. 8. It is observed that $K_1 = U/2$ gives the sharpest PEP metric but its sidelobe peaks are larger. It is worth-noting that minimizing the CRB [12] gives the same result of $K_1 = U/2$. The PEPs for $K_1 = U/2 + m$ and $K_1 = U/2 - m$ are observed to be the same.

In Fig. 9, the effects of different U and SNR_i on the PEP are depicted where $K_1 = U/2$ is used. A larger U or SNR_i results in a sharper PEP with smaller sidelobes. Hence, using a larger U , we can lower the PEP sidelobes in order to ensure accurate estimation at low SNR_i (due to fading) for a larger $d = kL/2$ where k is an odd positive integer. Similarly, if a larger estimation range is required, the sidelobe peaks associated with some appropriate d can be lowered by using a larger U (for example, the sidelobe peaks for $d = 2$ in Fig. 4 can be lowered by a larger U to achieve an estimation range of $|v| < 2/L = 1/4$).

In practical packet-based wireless communications, if the signal arriving at the receiver is in deep fade (below the receiver sensitivity) or the instantaneous signal to noise ratio is very low, the receiver will not be able to detect the signal. Hence, the frequency offset estimator's performance under such conditions is irrelevant to practical systems and typical mean-square error (MSE) performance measure may not reflect the exact performance for practical systems at low SNR. Therefore, we introduce a practical estimator performance measure named "practical MSE" which represents the mean-square estimation error given that the received signal power is above a threshold (related to the receiver sensitivity). We used $\text{SNR}_i = 0$ dB as our threshold for practical MSE.

Next, we address what sidelobe peak level of PEP (in other words, what value of U) gives a reliable result with an extended estimation range. Suppose the sidelobe peak of the PEP is p at $\text{SNR}_i = 0$ dB. Then the contribution of the sidelobe to the practical MSE is approximately $p/(4L^2)$ which should be less than the practical MSE for $d = 0$ (the reference MSE) to suppress the sidelobe problem. In design, the CRB at $\text{SNR}_i = 0$ dB for $d = 0$ (denoted by CRB_{ref}) can be used in place of the reference MSE. Recall that U for $d = 0$ is larger than U for $d > 0$ by one. The parameter U for $d > 0$ cases can be chosen such that the corresponding PEP sidelobe peak is less than $4L^2 \text{CRB}_{\text{ref}}$. For example, if CRB_{ref} is 4×10^{-8} and $L = 8$, the allowable p for the suppression of the sidelobe problem could be 10^{-5} or smaller from which U can easily be determined.

Following a similar approach from [12], we obtain the snapshot CRB of frequency offset estimation at a given SNR_i for an observation vector consisting of U sub-blocks with the

⁴The larger mean value of the likelihood metric for $d = 0$ is due to the fact that one more training sub-block is used in the estimation for $d = 0$.

corresponding time-indices $\{l_0, l_1, \dots, l_{U-1}\}$ as

$$\text{CRB}_{|\text{SNR}_i} = \frac{\text{SNR}_i^{-1}}{8\pi^2 L \left(\sum_{k=0}^{U-1} l_k^2 - \frac{(\sum_{n=0}^{U-1} l_n)^2}{U} \right)} \quad (26)$$

which is used to calculate CRB_{ref} . Note that the proposed design yields a smaller CRB than the conventional training structures. Proof of this fact and the derivation of CRB are given in Appendix-B.

A. Implementation

In the following, we discuss an implementation of the maximum likelihood frequency offset estimator for $d = (k_0 + n_1/n_2)L$ where k_0, n_1, n_2 are non-negative integers with $n_1 = 0$ if $n_2 = 1$, $n_1 \in \{1, \dots, n_2 - 1\}$ if $n_2 > 1$, and L/n_2 is an integer. Form a new vector β of length N_β as follows

$$\beta_n = \sum_{\{m,l\}:\theta_l(v) - \theta_m(v) = 2\pi v n L / n_2} \rho(m, l), \quad n = 0, 1, \dots, N_\beta - 1 \quad (27)$$

where β_n is the summation of all correlation terms $\{\rho(m, l)\}$ which have the same phase factor $(\theta_l(v) - \theta_m(v) = n(2\pi v L / n_2))$, (see (29)). If $k_0 = n_1 = 0$, then $N_\beta = U$ (if 10 training sub-blocks are transmitted, $U = 9$ in this case). If $n_1 = 0$ and $k_0 > 0$, then $N_\beta = U + k_0$ (if 10 training sub-blocks are transmitted, $U = 8$ in this case). If $n_1 > 0$, then $N_\beta = n_2(U + 1 + k_0) + n_1$ (if 10 training sub-blocks are transmitted, $U = 8$ in this case). Then the frequency offset estimate is given by

$$\hat{v} = \frac{n_2}{LN_{\text{fft}}} \arg \max_{-N_{\text{fft}}/2 \leq m < N_{\text{fft}}/2} \{\Re\{F_{N_{\text{fft}}}\beta\}\} \quad (28)$$

where $F_{N_{\text{fft}}}\beta$ gives the N_{fft} point FFT of β .

In the above implementation, the trial values of v are $\{\frac{m n_2}{LN_{\text{fft}}} : m = -\frac{N_{\text{fft}}}{2}, -\frac{N_{\text{fft}}}{2} + 1, \dots, \frac{N_{\text{fft}}}{2} - 1\}$. Note that $N_{\text{fft}} \geq 2N_\beta$ and N_{fft} is a power of 2 for low-complexity FFT implementation. If necessary, a quadratic interpolation can be applied to $[\hat{v} - \frac{n_2}{LN_{\text{fft}}}, \hat{v}, \hat{v} + \frac{n_2}{LN_{\text{fft}}}]$ to fine-tune the frequency estimate. A larger N_{fft} is associated with a larger complexity while giving a better estimation accuracy especially if the quadratic interpolation is not performed. With the quadratic interpolation, a suitable choice of N_{fft} for low complexity, while giving no noticeable degradation in the estimation accuracy for SNR of practical interest, would be $2^{\lceil \log_2(4N_\beta) \rceil}$ or $2^{\lceil \log_2(8N_\beta) \rceil}$. Note that for $k_0 = n_1 = 0$, this implementation is exactly the same as [9].

B. Complexity

The estimation computational complexities associated with the conventional and the proposed approaches are presented in Table I for a training signal consisting of $2K + 2$ identical sub-blocks. The parameter η accounts for the complexity reduction due to the zero inputs to the FFT [9]. The number of non-zero FFT inputs U' for the proposed approach depends on d and U' is at most $3K - 1$ and can be smaller than that. For $L = 8$, $U = 18$ (=19 for $d = 0$) and $N_{\text{fft}} = 256$ for $d = kL$ and $N_{\text{fft}} = 512$ for $d = kL/2$, the number of real multiplications for $d = kL/2$ is about 1.33 times, for $d = kL > 0$ is about 0.96 times that of $d = 0$ and the numbers of real additions

are for $d = kL/2$, 1.45 times and for $d = kL > 0$, 0.97 times that of $d = 0$. Recall that MLE#1 [9] has complexity about L times that of $d = 0$ and hence our proposed approach would be a better choice for the estimation range $|v| < 1/L$.

IV. SIMULATION RESULTS AND DISCUSSIONS

A multipath Rayleigh fading channel with $L = 8$ taps and an exponential power delay profile with a 3 dB per tap decay factor is considered. Channel gains are assumed to remain constant over the whole training signal. In order to have the same frequency resolution in the estimation, $N_{\text{fft}} = 256$ is used for $d = kL$ and $N_{\text{fft}} = 512$ is used for $d = (k + 0.5)L$ where k is a non-negative integer. To decouple the effect of FFT bin resolution in performance comparison, we set frequency offsets on the FFT grids which would give optimistic results but does not affect the performance comparison.

In Fig. 10, the MSEs and practical MSEs are presented for different values of the training signal separation distance d with a frequency offset $v = 3/64$ which is within the frequency offset estimation ranges for all d (all approaches). At high SNR, a larger d gives a better performance due to a sharper likelihood (or PEP) metric mainlobe. At low SNR, $d = 0$ has the smallest MSE due to the smallest sidelobe peaks of its likelihood (or PEP) metric. From Figs. 5-7, it can be observed that $d = mL/2$ has a smaller sidelobe peaks for an even integer m than an odd m (note that the estimation range for an even m is half of that for an odd m). This fact translates into a better MSE for an even m at low SNR in Fig. 10. At low SNR, practical MSE is smaller than MSE which indicates that the conventional MSE performance at low SNR is pessimistic. Some $d > 0$ can give a better practical MSE than $d = 0$ depending on the sidelobe peaks of the associated likelihood (or PEP) metric.

In Fig. 11, the MSEs and practical MSEs obtained with different d are presented for $v = 3/32$. Since v is larger than the estimation range of the existing approaches (corresponding to $d = kL$ with k being a non-negative integer), their corresponding MSEs or practical MSEs are very high. Due to the extended estimation range, the proposed approach using $d = (k + 0.5)L$ still gives a reliable estimate.

Next, we discuss how to circumvent the problem for $d > 0$ having a larger practical MSE than the conventional structure ($d = 0$) at low SNR. In fact, if we certainly know that the frequency offset is limited to $|v| < 1/(2L)$, the above sidelobe problem can easily be avoided for most of the cases by limiting the search range to $|v| < 1/(2L)$ since the larger sidelobe peaks of the likelihood (or PEP) metric are around $|v| = 1/(2L)$. That means the sidelobe problem associated with some d in Fig. 10 can be relieved by limiting the search range. If the frequency offset can be larger than $1/(2L)$ but less than $1/L$, then the search range cannot be limited as above. In this case, U can be increased to avoid this sidelobe problem as discussed in the previous section. From CRB and Fig. 9, we can easily check that $U = 8$ and 12 for $d = 20$ will encounter the sidelobe problem while $U = 18$ will avoid the problem. In Figs. 12 and 13, the MSE and practical MSE performance obtained with $U = 18$ are presented for $v = 3/64$

and $v = 3/32$, respectively. As expected, the sidelobe problem is avoided for practical MSE.

In Figs. 14-16, we present a comparison of the proposed method with different parameters and the existing methods from [5] and [9] in terms of the estimation MSE and the estimation range. The training signal for [5] is generated by 64-point IFFT of a length-64 Golay complementary sequence and then repeating it once and adding a cyclic prefix of 8 samples. Hence for [5], 136 training samples are transmitted and 128 training samples are used in the frequency offset estimation. The training signal for [9] is composed of 17 identical sub-blocks (including the CP) of 8 samples each (total 136 samples) and 16 sub-blocks (128 samples) are used in the estimation. For the proposed method, two sets of parameters are considered: “ $U = 15$ (17 identical sub-blocks of 8 samples each (total 136 samples) are transmitted), $K_1 = 8$ ”, and “ $U = 16$ (18 identical sub-blocks are transmitted), $K_1 = U/2 = 8$ ”. For each set, $d = 4$ and 12 are used. The estimation ranges of [5], [9], and the proposed method are $|v| < 1/128$, $|v| < 1/16$, and $|v| < 1/8$, respectively. From the simulation results in Figs. 14-16, we can also observe these estimation ranges. The estimation range of [5] is very small even compared with that of [9] and the proposed method’s estimation range is twice that of [9]. In terms of practical MSE performance, [9] has a better performance than [5] while the proposed method outperforms both [5] and [9].

V. CONCLUSIONS

For emerging high data-rate communication systems in highly dispersive channels such as ultra-wideband systems, possible frequency offsets could be larger than the estimation range of the existing methods using training signals with several identical sub-blocks. This paper addressed this issue and presented a maximum likelihood estimator with an extended estimation range as well as an improved estimation performance. The range extension and the estimation performance improvement are accomplished by designing the likelihood metric or a new design metric which is a pair-wise error probability (PEP) between the correct frequency offset point and a trial frequency offset point. The proposed new PEP metric gives more theoretical insights on the performance of practical maximum likelihood estimators. At comparable complexity with the same training overhead amount, the proposed method at least doubles the estimation range and also improves the estimation performance.

APPENDIX-A

In this appendix, we show that $Z(\tilde{v})$ and hence, $E[Z(\tilde{v})]/\text{Std}[Z(\tilde{v})]$ and PEP metrics just depend on $(\tilde{v} - v)$. By this fact, we can conclude that a change in v value will simply result in a shift in the metrics (in the \tilde{v} -axis) which does not affect the estimation range. Hence, in our design for estimation range extension, we can simply set $v = 0$.

Equation (15) can be expressed as

$$\begin{aligned} \rho(m, l) = & e^{j(\theta_l(v) - \theta_m(v))} \sum_{n=0}^{L-1} [|x(n)|^2 + x(n)\tilde{w}^*(mL+n) \\ & + x^*(n)\tilde{w}(lL+n) + \tilde{w}(lL+n)\tilde{w}^*(mL+n)] \end{aligned} \quad (29)$$

where $\{\tilde{w}(kL+n) = w(kL+n)e^{-j(\theta_k(v)+2\pi v n)}\}$ have the same statistics as $\{w(k)\}$. Similarly, the elements of $\Omega(v) - \Omega(\tilde{v})$ can be given by

$$\begin{aligned} & e^{j(\theta_m(v) - \theta_l(v))} - e^{j(\theta_m(\tilde{v}) - \theta_l(\tilde{v}))} \\ & = e^{j(\theta_m(v) - \theta_l(v))} \left(1 - e^{j(\theta_m(\tilde{v}-v) - \theta_l(\tilde{v}-v))} \right). \end{aligned} \quad (30)$$

By substituting (29) and (30) into (18), $Z(\tilde{v})$ can be expressed as

$$\begin{aligned} Z(\tilde{v}) & = \Re \left\{ \sum_{m=0}^{U-1} \sum_{l=m}^{U-1} \left[e^{j(\theta_m(v) - \theta_l(v))} - e^{j(\theta_m(\tilde{v}) - \theta_l(\tilde{v}))} \right] \rho(m, l) \right\} \\ & = \Re \left\{ \sum_{m=0}^{U-1} \sum_{l=m}^{U-1} \left(1 - e^{j(\theta_m(\tilde{v}-v) - \theta_l(\tilde{v}-v))} \right) \sum_{n=0}^{L-1} [|x(n)|^2 \right. \\ & \quad \left. + x(n)\tilde{w}^*(mL+n) + x^*(n)\tilde{w}(lL+n) \right. \\ & \quad \left. + \tilde{w}(lL+n)\tilde{w}^*(mL+n) \right\}. \end{aligned} \quad (31)$$

Equation (31) proves that the statistics of $Z(\tilde{v})$ just depend on $(\tilde{v} - v)$.

Next, we present a justification for the Gaussian approximation of $Z(\tilde{v})$. For SNR values of practical interest, the last term $\tilde{w}(lL+n)\tilde{w}^*(mL+n)$ from (31) is negligible if compared to the remaining terms. Hence, we can approximate $Z(\tilde{v})$ as

$$\begin{aligned} Z(\tilde{v}) \approx & \Re \left\{ \sum_{m=0}^{U-1} \sum_{l=m}^{U-1} \left(1 - e^{j(\theta_m(\tilde{v}-v) - \theta_l(\tilde{v}-v))} \right) \sum_{n=0}^{L-1} [|x(n)|^2 \right. \\ & \left. + x(n)\tilde{w}^*(mL+n) + x^*(n)\tilde{w}(lL+n) \right\}. \end{aligned} \quad (32)$$

For given S , \mathbf{h} and v , $\{x(n)\}$ are deterministic variables and $\{\tilde{w}(k)\}$ are Gaussian random variables. $Z(\tilde{v})$ in (32) is just a linear combination of Gaussian random variables and hence, it is a Gaussian random variable.

The insignificant term neglected in (32) can be expressed as

$$\begin{aligned} & \sum_{n=0}^{L-1} \left[\sum_{m=0}^{U-1} \sum_{l=m}^{U-1} \Re \left\{ \left(1 - e^{j(\theta_m(\tilde{v}-v) - \theta_l(\tilde{v}-v))} \right) \right. \right. \\ & \quad \left. \left. \times \tilde{w}(lL+n)\tilde{w}^*(mL+n) \right\} \right] = \sum_{n=0}^{L-1} \chi_n. \end{aligned} \quad (33)$$

Note that there are no common Gaussian noise terms in χ_n and χ_k for $n \neq k$. Since $\{\tilde{w}(l)\}$ are iid Gaussian random variables (with zero-mean, variance σ_n^2), $\{\chi_n\}$ are also iid random variables. This paper considers a highly dispersive channel and hence, the number of channel taps L is large. Then by the Central limit theorem, the term in (33) (and hence $Z(\tilde{v})$) can be approximated as a Gaussian random variable. In brief, for SNR of practical interest, the term from (33) can be neglected and in this case $Z(\tilde{v})$ is exactly a Gaussian random variable. If the (insignificant) term from (33) is included, $Z(\tilde{v})$ can be approximated as a Gaussian random variable by means of the Central limit theorem.

APPENDIX-B

In this appendix, we first derive the snap-shot CRB of the frequency offset estimation at a given SNR; for an observation vector consisting of U sub-blocks with the corresponding time-indices $\{l_0, l_1, \dots, l_{U-1}\}$. Then we prove that the proposed design gives a smaller CRB than the conventional

training structures consisting of several consecutive identical sub-blocks.

Following the same approach from [9], we obtain the snapshot CRB for a given \mathbf{h} as

$$\text{CRB}_{|\mathbf{h}} = \frac{\sigma_n^2}{2\mathbf{y}^H(\mathbf{I}_{UL} - \mathbf{B})\mathbf{y}} \quad (34)$$

where

$$\mathbf{y} = 2\pi\Psi\mathbf{S}\mathbf{h} \quad (35)$$

$$\Psi = \text{diag}\{l_0, l_0 + 1, \dots, l_0 + L - 1, \\ l_1, l_1 + 1, \dots, l_1 + L - 1, \dots, \\ l_{U-1}, l_{U-1} + 1, \dots, l_{U-1} + L - 1\}. \quad (36)$$

After some manipulation in (35) and substituting (13) into (34), we obtain

$$\mathbf{y} = 2\pi \left[(\Psi_0\mathbf{S}_0\mathbf{h} + l_0\mathbf{S}_0\mathbf{h})^T, (\Psi_0\mathbf{S}_0\mathbf{h} + l_1\mathbf{S}_0\mathbf{h})^T, \dots, (\Psi_0\mathbf{S}_0\mathbf{h} + l_{U-1}\mathbf{S}_0\mathbf{h})^T \right]^T \quad (37)$$

$$\mathbf{y}^H\mathbf{I}_{UL}\mathbf{y} = 4\pi^2 \left(U\mathbf{h}^H\mathbf{S}_0^H\Psi_0^2\mathbf{S}_0\mathbf{h} + 2\mathbf{h}^H\mathbf{S}_0^H\Psi_0\mathbf{S}_0\mathbf{h} \sum_{k=0}^{U-1} l_k + \sum_{n=0}^{U-1} l_n^2\mathbf{h}^H\mathbf{S}_0^H\mathbf{S}_0\mathbf{h} \right) \quad (38)$$

$$\mathbf{y}^H\mathbf{B}\mathbf{y} = 4\pi^2 \left(U\mathbf{h}^H\mathbf{S}_0^H\Psi_0^2\mathbf{S}_0\mathbf{h} + 2\mathbf{h}^H\mathbf{S}_0^H\Psi_0\mathbf{S}_0\mathbf{h} \sum_{k=0}^{U-1} l_k + \frac{\left(\sum_{n=0}^{U-1} l_n\right)^2}{U} \mathbf{h}^H\mathbf{S}_0^H\mathbf{S}_0\mathbf{h} \right) \quad (39)$$

where \mathbf{S}_0 is a $L \times L$ matrix with elements $[\mathbf{S}_0]_{k,n} = s_{(k-n)_L}$ and $\Psi_0 = \text{diag}\{0, 1, \dots, L-1\}$. Substituting the above equations into (34) together with the definitions $\mathbf{h}^H\mathbf{S}_0^H\mathbf{S}_0\mathbf{h} = E_1$ and $\text{SNR}_i = E_1/(L\sigma_n^2)$ gives the snap-shot CRB as in (26). The snap-shot CRB for a given \mathbf{h} depends on the snap-shot (instantaneous) SNR and hence, the notation $\text{CRB}_{\text{SNR}_i}$ has been used throughout the paper.

Define

$$\alpha = \sum_{k=0}^{U-1} l_k^2 - \frac{\left(\sum_{n=0}^{U-1} l_n\right)^2}{U}. \quad (40)$$

Then we have

$$\text{CRB}_{|\text{SNR}_i} = \frac{\text{SNR}_i^{-1}}{8\pi^2 L\alpha}. \quad (41)$$

In the following, we calculate α for the *conventional* structure denoted by α_c and for the *proposed* design denoted by α_p . In the proposed design, the training signal is separated into two groups of the same length and the two transmitted training signal groups are distanced by $d > 0$ samples. Suppose that $2K + 2$ training sub-blocks are transmitted. For the conventional structure, we have $U = 2K + 1$, $l_k = kL$ for $k = 0, 1, \dots, 2K$. After some simplification in (40), we obtain

$$\alpha_c = \frac{L^2}{3}K(2K^2 + 3K + 1). \quad (42)$$

For the proposed design, we have $U = 2K$, $l_k = kL$ for $k = 0, 1, \dots, K-1$ and $l_{K+n} = (n+K+1)L + d = (n+K+d')L$ for $n = 0, 1, \dots, K-1$ where $d' = 1 + d/L$. After some calculation, we obtain

$$\alpha_p = \frac{L^2}{6}K\{4K^2 - 6K + 2 + (2K + d' - 1)(3d' + 3)\}. \quad (43)$$

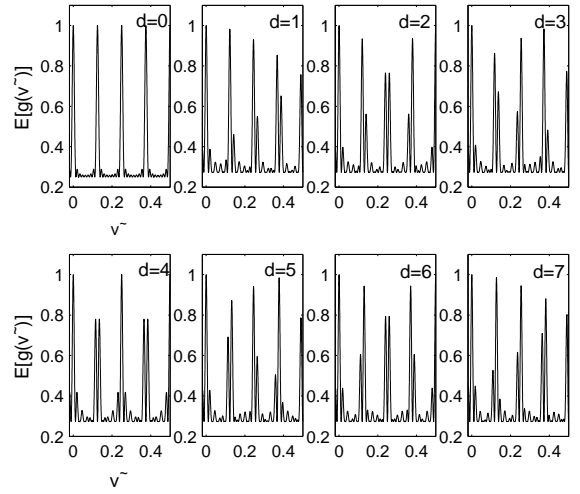


Fig. 2. The effects of the training signal group separation distance d on the normalized mean value of the likelihood metric for $d = 0, \dots, L-1$. The actual frequency offset corresponds to $\tilde{v} = 0$. Any other \tilde{v} with normalized $E[g(\tilde{v})]$ close to one causes an ambiguity in estimation. ($L = 8$)

Substituting $d' = 1$ into (43) yields $\alpha_p = \alpha_c$. Since $d' > 1$ and α_p increases with d' , we have proved that $\alpha_p > \alpha_c$ which means (see (41)) the proposed design gives a smaller snapshot CRB (and hence a smaller average CRB, averaged over the channel fading statistics) than the conventional training structure does.

REFERENCES

- [1] F. Classen and H. Meyr, "Frequency synchronization algorithms for OFDM systems suitable for communication over frequency selective fading channels," *IEEE VTC*, June 1994, pp. 1655-1659.
- [2] F. Daffara and O. Adami, "A new frequency detector for orthogonal multicarrier transmission techniques," *IEEE VTC*, July 1995, pp. 804-809.
- [3] T. M. Schmidl and D. C. Cox, "Robust frequency and timing synchronization for OFDM," *IEEE Trans. Commun.*, pp. 1613-1621, Dec. 1997.
- [4] T. Keller, L. Piazza, P. Mandarini, and L. Hanzo, "Orthogonal frequency division multiplex synchronization techniques for frequency-selective fading channels," *IEEE J-SAC*, pp. 999-1008, June 2001.
- [5] P.H. Moose, "A technique for orthogonal frequency division multiplexing frequency offset correction," *IEEE Trans. Commun.*, pp. 2908-2914, Oct. 1994.
- [6] J-J. van de Beek, M. Sandell and P.O. Börjesson, "ML estimation of time and frequency offset in OFDM systems," *IEEE Trans. Signal Proc.*, pp. 1800-1805, July 1997.
- [7] M.G. Hebley and D. P. Taylor, "The effect of diversity on a burst-mode carrier-frequency estimator in the frequency-selective multipath channel," *IEEE Trans. Commun.*, pp. 553-560, Apr. 1998.
- [8] M. Luise and R. Reggiannini, "Carrier frequency acquisition and tracking for OFDM systems," *IEEE Trans. Commun.*, pp. 1590-1598, Nov. 1996.
- [9] M. Morelli and U. Mengali, "Carrier-frequency estimation for transmissions over selective channels," *IEEE Trans. Commun.*, pp. 1580-1589, Sept. 2000.
- [10] H. Minn, P. Tarasak and V. K. Bhargava, "OFDM frequency offset estimation based on BLUE principle," *IEEE VTC (Fall)*, Sept. 2002, pp. 1230-1234.
- [11] —, "Some issues of complexity and training symbol design for OFDM frequency offset estimation methods based on BLUE principle," *IEEE VTC (Spring)*, Apr. 2003, pp. 1288-1292.
- [12] H. Minn, S. Xing, and V. K. Bhargava, "Optimal periodic training signal for frequency offset estimation in frequency selective fading channels," *IEEE ICC*, June 2004, pp. 488-492.
- [13] IEEE LAN/MAN Standards Committee, "Wireless LAN medium access control (MAC) and physical layer (PHY) specifications: High-speed physical layer in the 5 GHz band," *IEEE Standard 802.11a*, 1999.

TABLE I
ESTIMATION COMPUTATIONAL COMPLEXITY FOR A TRAINING SIGNAL CONSISTING OF $2K + 2$ IDENTICAL SUB-BLOCKS

# real multiplication	$2LU(U + 1) + 2N_{\text{fft}} \eta \log_2(N_{\text{fft}})$
# real addition	$2LU(U + 1) - 2 + 3N_{\text{fft}} \eta \log_2(N_{\text{fft}})$
	$\eta = 1 - \frac{\log_2(N_{\text{fft}}/U') + 2(U'/N_{\text{fft}} - 1)}{\log_2(N_{\text{fft}})}$
Conventional	$U' = 2K + 1$ $N_{\text{fft}} = N'$
Proposed	$U' = 3K - 1$ (at most) $N_{\text{fft}} = N'$ for $d = kL$ $N_{\text{fft}} = 2N'$ for $d = kL/2$

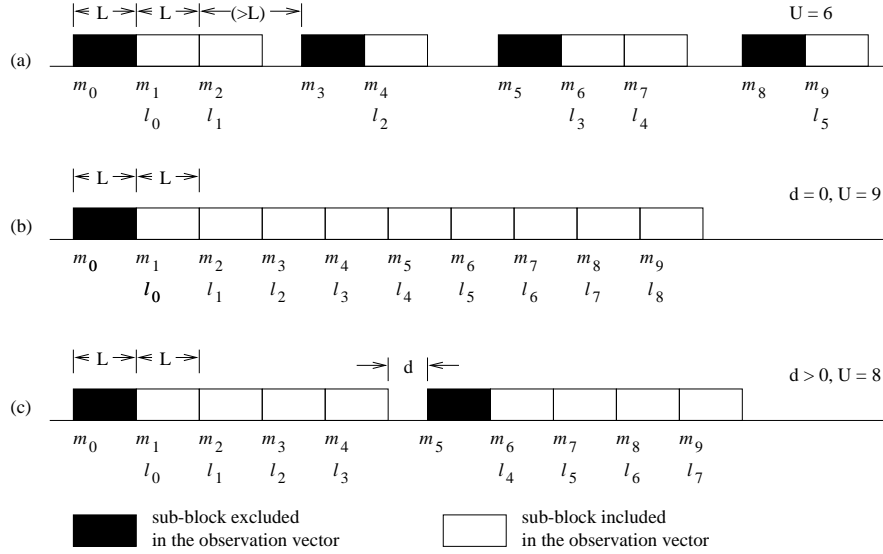


Fig. 1. Several training structures for $V = 10$ total transmitted identical training signal sub-blocks (defined by $\{m_k\}$) and the corresponding observation vector's sub-blocks (defined by $\{l_k\}$). (a) An arbitrary training signal structure, (b) The existing training signal structure (Our considered training signal structure with $d = 0$), (c) Our considered training signal structure with $d > 0$.

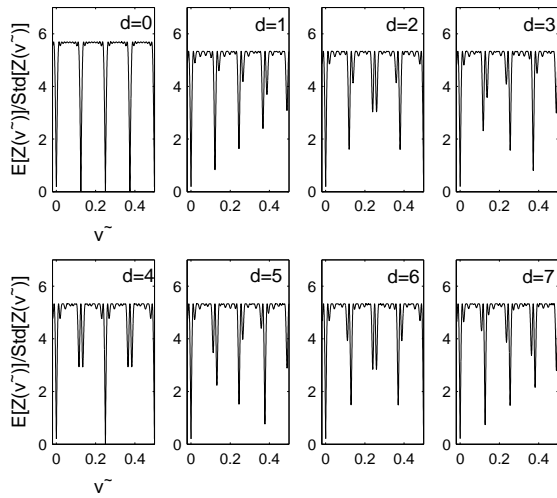


Fig. 3. The effects of the training signal group separation distance d on the ratio $E[Z(\tilde{v})]/\text{Std}[Z(\tilde{v})]$ for $d = 0, \dots, L - 1$. The actual frequency offset corresponds to $\tilde{v} = 0$. Any other \tilde{v} with $E[Z(\tilde{v})]/\text{Std}[Z(\tilde{v})]$ close to zero causes an ambiguity in estimation. ($L = 8$)

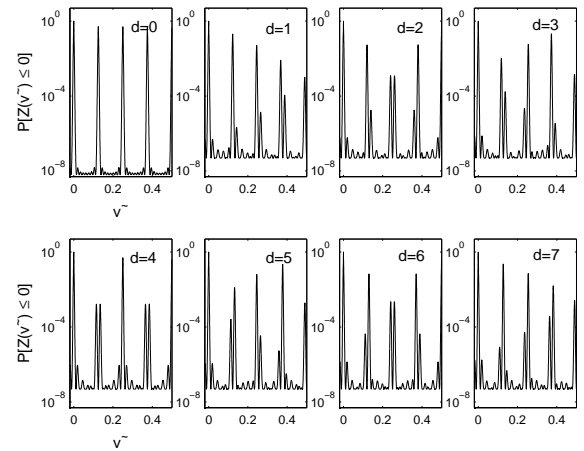


Fig. 4. The effects of the training signal group separation distance d on the pair-wise error probability (PEP) metric for $d = 0, \dots, L - 1$. The actual frequency offset corresponds to $\tilde{v} = 0$. Any other \tilde{v} with PEP close to 0.5 causes an ambiguity in estimation. ($L = 8$)

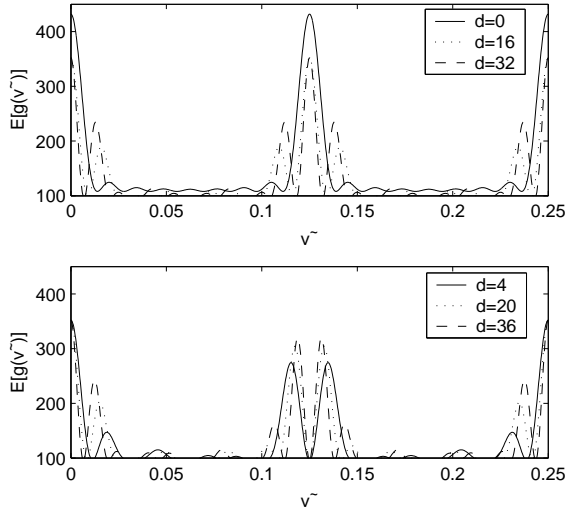


Fig. 5. The effects of the training signal group separation distance d on the mean value of likelihood metric for $d = kL/2$ where k is an integer. The actual frequency offset corresponds to $\tilde{v} = 0$. ($L = 8$)

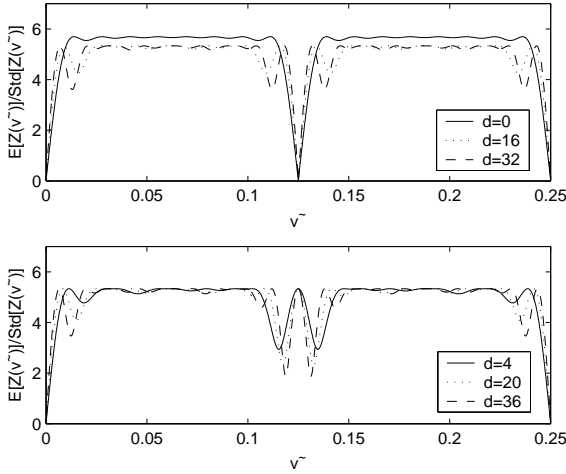


Fig. 6. The effects of the training signal group separation distance d on the ratio $E[Z(\tilde{v})]/\text{Std}[Z(\tilde{v})]$ for $d = kL/2$ where k is an integer. The actual frequency offset corresponds to $\tilde{v} = 0$. ($L = 8$)

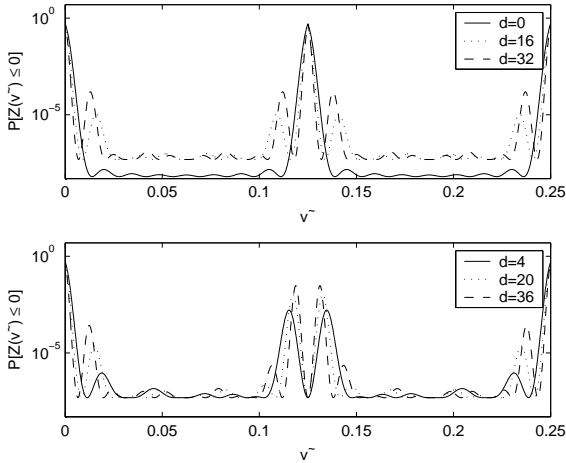


Fig. 7. The effects of the training signal group separation distance d on the pair-wise error probability metric for $d = kL/2$ where k is an integer. The actual frequency offset corresponds to $\tilde{v} = 0$. ($L = 8$)

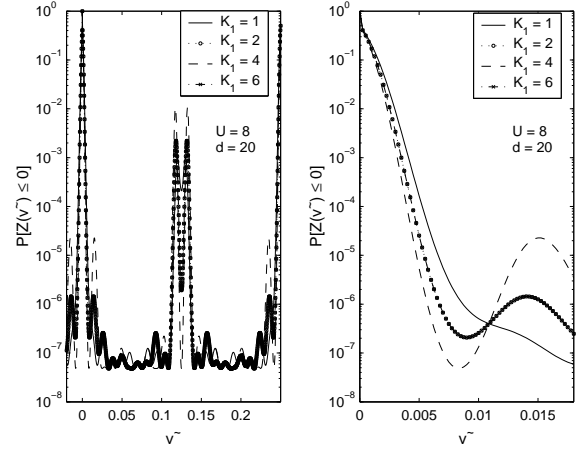


Fig. 8. The effects of K_1 (the number of training sub-blocks in the first group) on the pair-wise error probability metric (where total number of training sub-blocks U of the two groups is fixed). The actual frequency offset corresponds to $\tilde{v} = 0$. ($L = 8$)

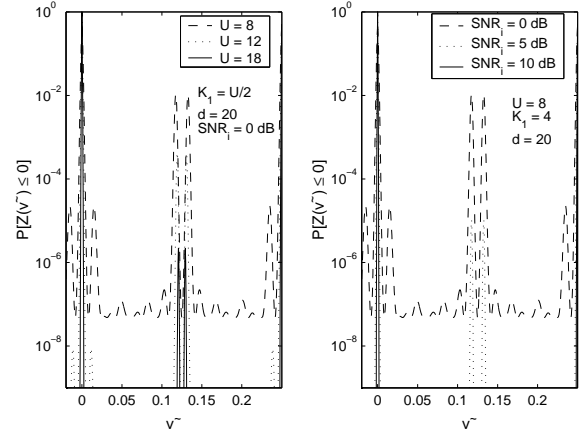


Fig. 9. The effects of total number of training sub-blocks U and SNR_i on the pair-wise error probability metric. The actual frequency offset corresponds to $\tilde{v} = 0$. ($L = 8$)

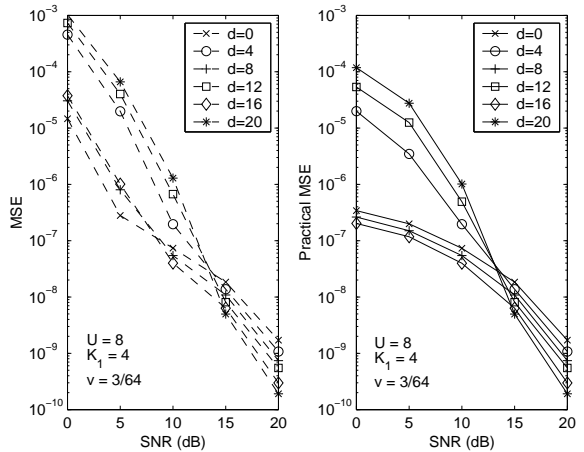


Fig. 10. Estimation performance obtained with different training separation distance d for a normalized frequency offset v less than $1/(2L)$. ($L = 8$)

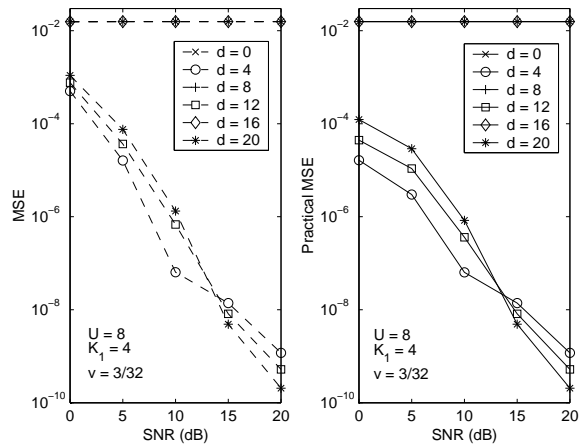


Fig. 11. Estimation performance obtained with different training separation distance d for a normalized frequency offset v between $1/(2L)$ and $1/L$. ($L = 8$)

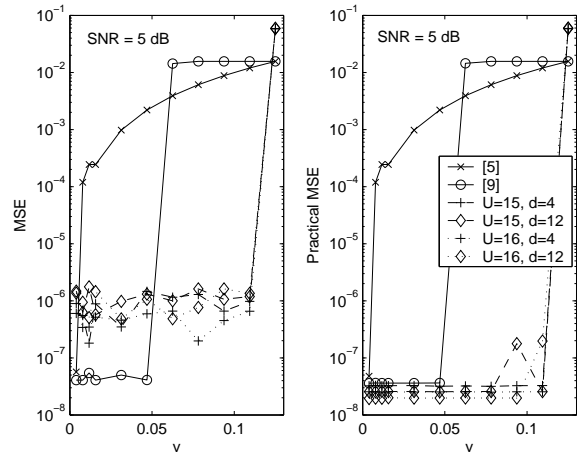


Fig. 14. Comparison of estimation MSE and estimation range for several methods at SNR = 5 dB

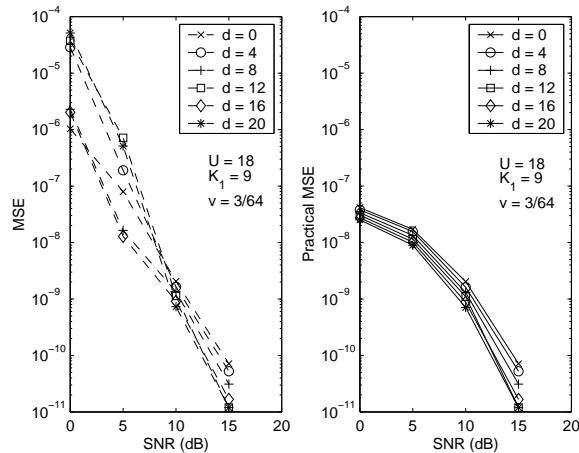


Fig. 12. Estimation performance obtained with a larger U (for suppressing sidelobe peaks) for a normalized frequency offset v less than $1/(2L)$. ($L = 8$)

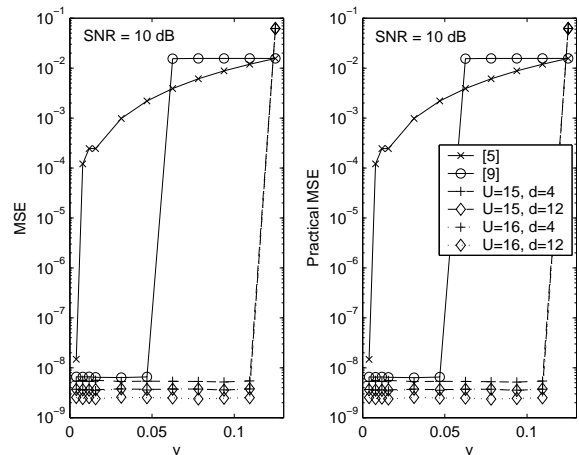


Fig. 15. Comparison of estimation MSE and estimation range for several methods at SNR = 10 dB

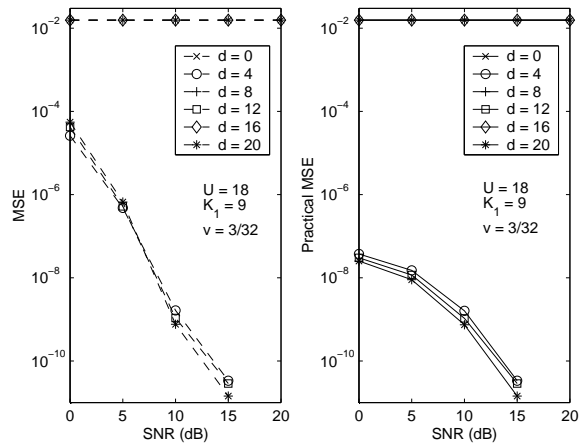


Fig. 13. Estimation performance obtained with a larger U (for suppressing sidelobe peaks) for a normalized frequency offset v between $1/(2L)$ and $1/L$. ($L = 8$)

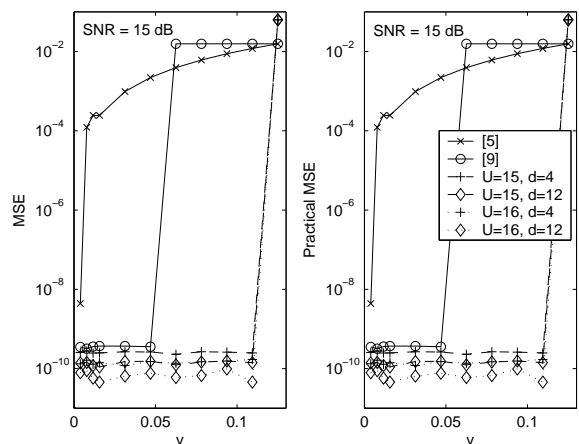


Fig. 16. Comparison of estimation MSE and estimation range for several methods at SNR = 15 dB