Optimal Periodic Training Signal for Frequency Offset Estimation in Frequency-Selective Fading Channels

Hlaing Minn, *Member, IEEE*, Xiaoyu Fu, *Student Member, IEEE*, and Vijay K. Bhargava, *Fellow, IEEE*

***Abstract*—This paper addresses an optimal periodic training signal design for frequency offset estimation in frequency-selective multipath Rayleigh fading channels. For a fixed transmitted training signal energy within a fixed-length block, the optimal periodic training signal structure (the optimal locations of identical training subblocks) and the optimal training subblock signal are presented. The optimality is based on the minimum Cramer–Rao bound (CRB) criterion. Based on the CRB for joint estimation of frequency offset and channel, the optimal periodic training structure (optimality only in frequency offset estimation, not necessarily in joint frequency offset and channel estimation) is derived. The optimal training subblock signal is obtained by using the average CRB (averaged over the channel fading) and the received training signal statistics. A robust training structure design is also presented in order to reduce the occurrence of outliers at low signal-to-noise ratio values. The proposed training structures and subblock signals achieve substantial performance improvement.**

***Index Terms*—Cramer–Rao bound (CRB), frequency offset estimation, training signal design, training structure, zero autocorrelation (ZAC).**

I. Introduction

TRAINING signals are commonly used in communications systems for timing synchronization, frequency synchronization, and channel estimation. Training signal design is an important issue, since a proper design can significantly improve the estimation performance and can drastically reduce the estimation complexity.

In [1], periodic training sequences for channel equalization were addressed, and zero autocorrelation (ZAC) sequences were shown to be optimal for the cyclic convolution-type channel equalization. Efficient computer search methods for the best periodic and aperiodic training sequences in least-squares channel estimation were presented in [2]–[4] for several channel lengths and training sequence lengths. Constructions of optimal complex sequences were discussed in [5]–[8] and references therein. Recent treatments of the optimal training structure and training signal design for channel estimation can be found in [9]–[11] and references therein.

Training sequence design for timing synchronization was discussed in [12] in the context of the minimum-shift keying signal. For timing synchronization, [13] presented the best patterns $\{p_k \in \pm 1\}$ for the training signal of the type $[p_1\boldsymbol{A}, p_2\boldsymbol{A}, \ldots]$ where $\pm\boldsymbol{A}$ represents a training subblock signal.

Regarding training design for frequency synchronization, [14] addressed optimal training signals in additive white Gaussian noise (AWGN) channels by minimizing the Cramer–Rao bound (CRB). In [15], an optimal training signal design for frequency offset estimation when the channel impulse response (CIR) is perfectly known was presented. The resulting training signal depends on the CIR. Optimal training signal design for frequency offset estimation in frequency-selective channels is not an easy task. This problem was addressed in [16] by applying a minmax approach based on an asymptotic CRB. The channel gains remain constant within the training block, and the asymptotic CRB is obtained by setting the training block length (in samples) $N \to \infty$. For a fixed channel energy, the minmax approach minimizes the asymptotic CRB for the worst-case channel response. The optimality of the training signal from [16] is limited by the minmax approach and the asymptotic CRB, but [16] does provide a neat solution to a long-standing problem. It is also noted that due to the fixed channel energy constraint, the channel fading effect is not included in the training signal design of [16]. Hence, the optimal training signal design for frequency offset estimation in frequency-selective fading channels is still an open problem, which will be addressed in this paper.

Recently, in [17], the authors of [16] nicely extended their minmax approach in minimizing the asymptotic CRB to the problem of training signal design for joint estimation of frequency offset and CIR. The estimation errors of frequency offset and CIR are equally weighted in their design criterion. Since the estimation error effects of frequency offset and CIR on the bit-error rate (BER) may not be equal, finding the optimal weighting between estimation errors of frequency offset and CIR and designing the corresponding optimal training signal still constitute a delicate open problem. In this paper, we focus on the training signal design for the frequency offset estimation. This gives the best design for the systems which use separate training signals for estimation of frequency offset and CIR (in receivers with low complexity, small delay, and small buffer size). We also

Paper approved by J. Wang, the Editor for Wireless Spread Spectrum of the IEEE Communications Society. Manuscript received June 25, 2004; revised April 24, 2005 and October 28, 2005. This work was supported in part by the School of Engineering and Computer Science, University of Texas at Dallas, and in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada. This paper was presented in part at the IEEE International Conference on Communications, 2004.

H. Minn and X. Fu are with the Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX 75083-0688 USA (e-mail: hlaing.minn@utdallas.edu; xxf031000@utdallas.edu).

V. K. Bhargava is with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: vijayb@ece.ubc.ca).

Digital Object Identifier 10.1109/TCOMM.2006.876869

0090-6778/$20.00 © 2006 IEEE

evaluate, by simulation, the error performance using the same proposed training signal for estimation of both frequency offset and CIR, and present the advantages of our proposed training signal.

For the ease of estimation complexity, we consider periodic training signals consisting of several identical subblocks. Periodic training signals have been extensively used in practice (e.g., GSM, IEEE 802.11a, 802.15, 802.16). In this paper, we will answer the following question: "For a fixed transmitted training signal energy, what is the optimal periodic training structure (the locations of the training subblocks) within a fixed-length block, and what is the optimal training signal (of a subblock) that gives the minimum CRB of the frequency offset estimation in frequency-selective multipath Rayleigh fading channels?" We will address the question in two steps. In the first step, we will find the optimal training structure for a fixed block length and a fixed *received* training signal energy. In the second step, we will investigate the optimal training signal (of a subblock) for a fixed *transmitted* training signal energy in a frequency-selective multipath Rayleigh fading channel. The combination of the results from the two steps will give the solution to the above question. Also note that although the CRB of the frequency offset estimation adopted in this paper is based on the joint estimation of frequency offset and CIR, we only consider training design for frequency offset estimation, and hence, the optimality is only for frequency offset estimation, not necessarily for joint estimation of frequency offset and CIR.

The rest of the paper is organized as follows. Section II describes the signal model and the CRB. Section III presents the optimal periodic training structure. In Section IV, the optimal training signal (of a subblock) is addressed. Simulation results and discussions are provided in Section V. Finally, conclusions are given in Section VI.

II. Signal Model and CRB

We consider a wide-sense-stationary uncorrelated-scattering frequency-selective Rayleigh fading channel characterized by L uncorrelated complex baseband tap gains $h(0), h(1), \ldots, h(L-1)$ and tap spacing of symbol duration T_s. The channel is assumed to be quasi-static, where the tap gains remain essentially constant over the block length NT_s. The complex baseband received signal sampled at the symbol rate can be expressed as

$$r(n) = x(n)e^{j2\pi nv} + w(n), \quad n = 0, 1, \ldots, N-1 \tag{1}$$

where v is the carrier frequency offset normalized by the symbol rate $1/T_s$, $\{w(n)\}$ are the uncorrelated zero-mean circularly symmetric complex Gaussian noise samples, each having a variance of σ_n^2, and $\{x(n)\}$ are the channel output signal samples which are related to the transmitted signal samples[1] $\{s_n : n = -L, -L+1, \ldots, N-1\}$ as

$$x(n) = \sum_{k=0}^{L-1} h(k)s_{n-k}, \quad n = 0, 1, \ldots, N-1. \tag{2}$$

[1]Although we do not need $s(-L)$ to obtain the intersymbol interference (ISI)-free received samples in (1), $s(-L)$ is included for convenience and consistency with the considered context of the periodic training signal consisting of $(P+1)$ identical subblocks of length L samples each. The ISI-affected received samples $r(-L), \ldots, r(-1)$ are discarded.

In matrix form, the received signal vector is expressed as

$$\boldsymbol{r} = \boldsymbol{W}(v)\boldsymbol{S}\boldsymbol{h} + \boldsymbol{w} = \boldsymbol{W}(v)\boldsymbol{x} + \boldsymbol{w} \tag{3}$$

where

$$\boldsymbol{r} = [r(0), r(1), \ldots, r(N-1)]^T \tag{4}$$
$$\boldsymbol{h} = [h(0), h(1), \ldots, h(L-1)]^T \tag{5}$$
$$\boldsymbol{x} = [x(0), x(1), \ldots, x(N-1)]^T \tag{6}$$
$$\boldsymbol{w} = [w(0), w(1), \ldots, w(N-1)]^T \tag{7}$$
$$\boldsymbol{W}(v) = \mathrm{diag}\left\{1, e^{j2\pi v}, \ldots, e^{j2\pi(N-1)v}\right\} \tag{8}$$
$$[\boldsymbol{S}]_{ij} = s_{i-j}, \ \ 0 \le i \le N-1, \ \ 0 \le j \le L-1 \tag{9}$$

and $\mathrm{diag}\{\}$ represents a diagonal matrix of the specified diagonal elements. The covariance matrices for $\boldsymbol{w}$ and $\boldsymbol{h}$ are given by $\boldsymbol{C_w} = E[\boldsymbol{w}\boldsymbol{w}^H] = \sigma_n^2\boldsymbol{I}_N$ and $\boldsymbol{C_h} = E[\boldsymbol{h}\boldsymbol{h}^H] = \mathrm{diag}\{\sigma_{h_0}^2, \sigma_{h_1}^2, \ldots, \sigma_{h_{L-1}}^2\}$, where trace $(\boldsymbol{C_h})$ is set to unity and $\boldsymbol{I}_N$ is the $N \times N$ identity matrix. The superscripts T and H represent the transpose and Hermitian transpose, respectively.

If $\{s_k\}$ are known training signal samples, then the CRB for v derived for the joint estimation of v and $\boldsymbol{h}$ based on the received vector $\boldsymbol{r}$ is given by [18]

$$\mathrm{CRB}_{|\boldsymbol{h}} = \frac{\sigma_n^2}{8\pi^2\boldsymbol{h}^H\boldsymbol{S}^H\boldsymbol{\Lambda}(\boldsymbol{I_N} - \boldsymbol{B})\boldsymbol{\Lambda}\boldsymbol{S}\boldsymbol{h}} \tag{10}$$

where

$$\boldsymbol{\Lambda} = \mathrm{diag}\{0, 1, \ldots, N-1\} \tag{11}$$
$$\boldsymbol{B} = \boldsymbol{S}(\boldsymbol{S}^H\boldsymbol{S})^{-1}\boldsymbol{S}^H. \tag{12}$$

Note that we have used the notation $\mathrm{CRB}_{|\boldsymbol{h}}$ to reflect that the CRB is just for a given channel realization $\boldsymbol{h}$. We will call it a snapshot CRB in this paper, since it corresponds to a snapshot channel realization.

If the whole training block of length $N+L = (P+1)L$ samples is constructed by repeating the same training signal subblock of length L samples, then $\boldsymbol{S}$ is obtained by P-fold stacking the submatrix $\boldsymbol{S}_o$ as

$$\boldsymbol{S} = \boldsymbol{1}_P \otimes \boldsymbol{S}_o \tag{13}$$

where $\otimes$ is the Kronecker product, $\boldsymbol{1}_P$ is the all-one column vector of length P, and $\boldsymbol{S}_o$ is an $L \times L$ circulant matrix with elements

$$[\boldsymbol{S}_o]_{i,j} = s_{|i-j|_L}, \quad 0 \le i, j \le L-1 \tag{14}$$

and $|i-j|_L$ means $(i-j)$ modulo L. The snapshot CRB for the periodic training signal with an arbitrary circulant $\boldsymbol{S}_o$ is derived in Appendix A. For a full-rank $\boldsymbol{S}_o$, the snapshot CRB can be simplified as

$$\mathrm{CRB}_{|\boldsymbol{h}} = \frac{3\mathrm{SNR}_{ss}^{-1}}{2\pi^2(PL^3)(P^2-1)} = \mathrm{CRB}_{|\mathrm{SNR}_{ss}} \tag{15}$$

where SNR_{ss} denotes the snapshot signal-to-noise ratio (SNR) and is defined by

$$\text{SNR}_{ss} = \frac{\boldsymbol{h}^H \boldsymbol{S}^H \boldsymbol{S}\boldsymbol{h}}{N\sigma_n^2}. \tag{16}$$

The snapshot CRB depends only on SNR_{ss} for fixed values of P and L. Hence, $\text{CRB}_{|\boldsymbol{h}}$ can be considered as $\text{CRB}_{|\text{SNR}_{ss}}$, the CRB conditioned on the snapshot SNR of the received training signal, as shown in (15).

III. Optimal Periodic Training Structure

Periodic training signals are commonly used in practice due to their advantage of complexity reduction in estimation. For example, using a periodic training signal block, which consists of $(P+1)$ subblocks of length L samples each, rather than a nonperiodic training signal block of the same length $N+L = (P+1)L$, would reduce the estimation complexity approximately by a factor of L [18]. For the same reason, we consider periodic training signals in this paper. Our interest is to find an optimal periodic training structure that gives the minimum (snapshot) CRB for a fixed L.

The problem is formulated as follows. "For a block with a fixed length of $(N+L)$ samples, which may contain training signals only or both training and data signals, and for a fixed received training signal energy $\mathcal{E}$, what is the best periodic training signal structure?" This problem is divided into two subproblems. In the first subproblem, we will investigate the best structure consisting of V identical training signal subblocks of length L, i.e., we have to find the best locations of $V(\leq P+1)$ identical training subblocks within the block of length $N+L = (P+1)L$ samples. In the second subproblem, we will find the best value of V.

Note that the frequency offset estimation is based on all training subblocks which are identical at the channel output (before being affected by the noise and frequency offset). Nonidentical channel output training subblocks are discarded. Let U be the number of training subblocks used in the estimation. If all V subblocks are consecutively located as one group, the first training subblock is not preceded by a training subblock, while each of the remaining training subblocks is. Then the first channel-output subblock is not identical to the other channel-output training subblocks and is discarded. So, we have $U = V-1$. If V subblocks form two groups of consecutive subblocks, where data or null signals are located between the two groups, then $U = V-2$, since the first subblock of each group is discarded in the estimation. In general, if there are G groups of consecutive training subblocks, we have $U = V-G$. Hence, forming more groups of consecutive training subblocks (a group may contain one or more consecutive subblocks) will result in a larger loss of training energy used in the estimation.

Fig. 1 depicts a structure consisting of several identical training subblocks and data or null subblocks. The indexes of transmitted training subblocks in the figure are $(-1, 0, 1), \ldots,$ and $(P-3, P-2)$. The subblocks with indexes $-1, \ldots,$ and $P-3$ are discarded, and hence, the observation vector contains

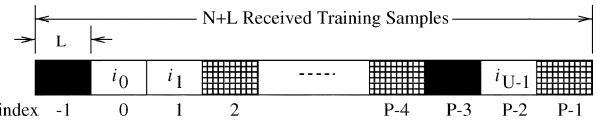


Fig. 1. Arbitrary periodic training structure illustrating the subblocks with indexes $[i_0, i_1, \ldots, i_{U-1}]$ used in the estimation. (The patterned subblocks are data or null signals, and the remaining subblocks are the transmitted training subblocks. The dark-shaded subblocks are not included in the observation vector of the estimation.)

the subblocks with indexes (0, 1), ..., and $P-2$. So, a general training structure can be defined by the observation vector's subblock location index vector $\boldsymbol{J} = [i_0, i_1, \ldots, i_{U-1}]$. The subblock indexes within the block are from -1 to $P-1$ and hence, in general, we have $i_0 \geq 0$, $i_{U-1} \leq P-1$, and $i_k < i_m$ for $k < m$. The observation vector can be given by

$$\boldsymbol{y} = \boldsymbol{Q}^T \boldsymbol{r} = \boldsymbol{\Gamma}(v)\boldsymbol{C}\boldsymbol{h} + \boldsymbol{Q}^T \boldsymbol{w} \tag{17}$$

where

$$\boldsymbol{\Gamma}(v) = \boldsymbol{Q}^T \boldsymbol{W}(v) \boldsymbol{Q} \tag{18}$$

$$\boldsymbol{Q} = \left[\boldsymbol{\mathcal{Q}}_{i_0}, \boldsymbol{\mathcal{Q}}_{i_1}, \ldots, \boldsymbol{\mathcal{Q}}_{i_{U-1}}\right] \tag{19}$$

$$\boldsymbol{\mathcal{Q}}_{i_k} = [\boldsymbol{e}_{Li_k}, \boldsymbol{e}_{Li_k+1}, \ldots, \boldsymbol{e}_{Li_k+L-1}] \tag{20}$$

$$\boldsymbol{C} = \boldsymbol{1}_U \otimes \boldsymbol{S}_o \tag{21}$$

and $\boldsymbol{e}_j$ is the $(j+1)$th column of $\boldsymbol{I}_N$. Note that $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}_{UL}$ and $\boldsymbol{S}_o$ is given by (14), with $\{s_k : k = 0, \ldots, L-1\}$ representing one training subblock.

The corresponding snapshot CRB can be given by (see Appendix A for details)

$$\text{CRB}_{|\text{SNR}_{ss}}(\boldsymbol{J}) = \frac{\text{SNR}_{ss}^{-1}}{8\pi^2 L^3 \left(\sum_{k=0}^{U-1} i_k^2 - \frac{\left(\sum_{n=0}^{U-1} i_n\right)^2}{U}\right)}. \tag{22}$$

For a general training structure consisting of $V = 2(K+1)$ identical training subblocks, the best training structure is determined by

$$\boldsymbol{J}_{2K+2}^* = \arg\min_{\boldsymbol{J}_{2K+2}} \text{CRB}_{|\text{SNR}_{ss}}(\boldsymbol{J}_{2K+2}) \tag{23}$$

where the vector $\boldsymbol{J}_{2K+2}$ is of variable length U, $0 \leq U \leq (2K+1)$, and it contains the location indexes of the U subblocks used in the estimation. Note that for each fixed value of U, there are several different $\boldsymbol{J}_{2K+2}$'s corresponding to different location index vectors. In the following, the solution to (23) will be pursued.

In [14], the optimal training signal for frequency offset estimation in an AWGN channel with an arbitrary phase shift was presented for the constraints of a fixed total training signal energy E_s, a fixed block length of N (samples), and a peak training sample energy P_ζ. For a training signal vector $\boldsymbol{s} = [s_0, s_1, \ldots, s_{N-1}]^T$ with $\boldsymbol{s}^H\boldsymbol{s} = E_s$, the CRB for frequency

offset estimation in the AWGN channel with an arbitrary phase shift is given by

$$\text{CRB} = \frac{N^2\sigma_n^2}{8\pi^2}\frac{1}{\boldsymbol{s}^H\boldsymbol{\Lambda}^2\boldsymbol{s} - \frac{(\boldsymbol{s}^H\boldsymbol{\Lambda}\boldsymbol{s})^2}{\boldsymbol{s}^H\boldsymbol{s}}} \tag{24}$$

and the optimal training signal that minimizes the above CRB while satisfying the peak training sample energy constraint ($\zeta_k \equiv |s_k|^2 \le P_\zeta$) is determined by the following optimal training sample energy allocation [14]:

$$\zeta_k^* = \begin{cases} \zeta_{N-1-k}^* = \min\left\{P_\zeta, E_s/2 - \sum_{l=0}^{k-1}\zeta_l^*\right\}, & \text{if } k=0,\ldots,\lfloor N/2\rfloor - 1 \\ E_s - 2\sum_{l=0}^{k-1}\zeta_l^*, & \text{if } k=(N-1)/2 \text{ and } N \text{ is an odd integer.} \end{cases} \tag{25}$$

Define

$$\tilde{\boldsymbol{s}} = [\tilde{s}_0, \tilde{s}_1, \ldots, \tilde{s}_{P-1}]^T \tag{26}$$

$$\tilde{s}_k = \begin{cases} 1, & \text{if } k \in \{i_0, \ldots, i_{U-1}\} \\ 0, & \text{otherwise} \end{cases} \tag{27}$$

$$\tilde{\boldsymbol{\Lambda}} = \text{diag}\{0, 1, \ldots, P-1\}. \tag{28}$$

Then (22) can be expressed as

$$\text{CRB}_{|\text{SNR}_{ss}}(\boldsymbol{J}) = \frac{\text{SNR}_{ss}^{-1}}{8\pi^2 L^3\left(\tilde{\boldsymbol{s}}^H\tilde{\boldsymbol{\Lambda}}^2\tilde{\boldsymbol{s}} - \frac{(\tilde{\boldsymbol{s}}^H\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{s}})^2}{\tilde{\boldsymbol{s}}^H\tilde{\boldsymbol{s}}}\right)}. \tag{29}$$

Hence, finding the best subblock indexes becomes finding $\tilde{\boldsymbol{s}}$ that minimizes (29). By setting $P_\zeta = 1$, $N = P$, and $E_s = U$ in (24), we observe that (29) and (24) have exactly the same form, except that (29) has an additional constraint of $\tilde{s}_k \in \{0,1\}$. The search range of $\tilde{\boldsymbol{s}}$ is just a subset of the search range of $\boldsymbol{s}$ in minimizing the corresponding CRBs. Since the optimal solution of $\boldsymbol{s}$ given by (25) is within the search range of $\tilde{\boldsymbol{s}}$, it is also the optimal solution of $\tilde{\boldsymbol{s}}$ for a fixed even U, where $1 \le U \le 2K$. Similarly, for a fixed odd U, where $1 \le U < 2K$, the optimal locations of the $U-1$ subblocks are given by (25), and the remaining subblock location is either $(U-1)/2$ or $P-(U+1)/2$ (c.f. [14, App. A]).[2] Also note that if the remaining subblock is not located at either $(U-1)/2$ or $P-(U+1)/2$, it cannot be used in the estimation. We simply skip $U = 0$ since it corresponds to the case where no training signal is used in the estimation. By substituting the optimal subblock location indexes into (22), we obtain that the snapshot CRB is smaller for a larger even U, and the snapshot CRB for $U = 2m$ is smaller than that for $U = 2m-1$, where m is an integer and $1 \le m \le K$. Consequently, the solution for $U = 2K$ gives the smallest snapshot CRB within $0 \le U \le 2K$. Hence, the best training structure for $U \le 2K$ and $P \ge 2K+2$ is given by its subblock location index vector as follows:[3]

$$\boldsymbol{J}_{2K+2}^* = [0, 1, \ldots, K-1, P-K, P-K+1, \ldots, P-1]. \tag{30}$$

The corresponding snapshot CRB is obtained by substituting (30) into (22) as

$$\text{CRB}_{|\text{SNR}_{ss}}\left(\boldsymbol{J}_{2K+2}^*\right) = \frac{3\text{SNR}_{ss}^{-1}}{4K\pi^2L^3(4K^2+3P^2-6PK-1)}. \tag{31}$$

The only possible value left to check is $U = (2K+1)$, which corresponds to the structure with consecutively located subblocks whose index vector is $\boldsymbol{J}_{2K+2} = \{i_k = k : k = 0, \ldots, 2K\}$. The snapshot CRB for $U = (2K+1)$ is given by (15), with P replaced by $2K+1$ as

$$\text{CRB}_{|\text{SNR}_{ss}}(\boldsymbol{J}_{2K+2}) = \frac{3\text{SNR}_{ss}^{-1}}{2\pi^2L^3(2K+1)\left((2K+1)^2-1\right)}. \tag{32}$$

A straightforward comparison between (31) and (32), together with $P \ge 2K+2$ for (31), yields that the snapshot CRB for $U = (2K+1)$ is larger than that for $\boldsymbol{J}_{2K+2}^*$. Hence, for $V = (2K+2)$, the best training structure is given by $\boldsymbol{J}_{2K+2}^*$ in (30).

If the block contains an odd number of training subblocks, say $V = (2K+3)$ subblocks, applying a similar approach for the case of an odd U and $V = (2K+2)$, we obtain the best location index vectors for $U \le (2K+1)$ and $P \ge 2K+3$ as follows:[4]

$$\boldsymbol{J}_{2K+3}^* = [0, 1, \ldots, K, P-K, P-K+1, \ldots, P-1] \tag{33}$$

$$\boldsymbol{J}_{2K+3}^* = [0, 1, \ldots, K-1, P-K-1, P-K, \ldots, P-1]. \tag{34}$$

A direct calculation after substituting the indexes from (33) and (34) into (22) shows that both structures give the same performance. The only possible value left is $U = (2K+2)$, which is the structure with consecutively located subblocks, and its snapshot CRB is given by (15) with P replaced by $2K+2$. By a direct comparison between the corresponding snapshot CRBs, we find that the structure in (33) or (34) has a smaller snapshot CRB than the structure for $U = (2K+2)$. Hence, both $\boldsymbol{J}_{2K+3}^*$'s represent the best location index vectors for the $V = 2K+3$ case.

Now, we investigate what value of V gives minimum snapshot CRB. We observe that there always exists a training structure with an even number of training signal subblocks which gives a smaller snapshot CRB than the structure with an odd number of training signal subblocks (see Appendix B for details). Hence, it is sufficient to consider an even number of training subblocks, i.e., $V = (2K+2)$.

[2]It shows that changing an arbitrary training structure into the structure of the type (30), (33), or (34), where K is replaced by U for the considered situation, always decreases the CRB in (24), and hence, the snapshot CRB in (29).

[3]Note that when $P = 2K+1$ and $U = 2K$, we do not need to discard the subblock with index $(P-K-1)$ in (30) since all subblocks are now consecutively located, and thus we have $U = 2K+1$ instead of $U = 2K$. In this case, the proposed structure is the same as the conventional structure.

[4]For $P = 2K+2$, we do not need to discard the subblock with index $(P-K-1)$ in (33) or $(P-K-2)$ in (34), and consequently, the proposed structure becomes the conventional structure with $U = P = 2K+2$.

Next, we investigate what value of K is the best for a block with a fixed length of $(P+1)L$ samples, a fixed total received training energy $\mathcal{E}$, and a fixed noise variance σ_n^2. Equation (31) can be expressed as

$$\mathrm{CRB}_{|\mathrm{SNR}_{ss}}\left(\boldsymbol{J}_{2K+2}^{*}\right)=\frac{3\sigma_n^2/\mathcal{E}}{2L^2\pi^2}f(K) \tag{35}$$

where

$$f(K)=\frac{K+1}{4K^3-6PK^2+(3P^2-1)K}. \tag{36}$$

The best value of K is determined by

$$K^{*}=\underset{1\le K\le\lceil (P-1)/2\rceil}{\arg\min} f(K) \tag{37}$$

which can easily be obtained by numerical evaluation. An analytical solution for $K^{\star}$ is presented in Appendix C. Note that since $f(K)$ is independent of SNR_{ss}, the best value K^{*} holds for any SNR_{ss} and hence, for fading channels as well. The optimal periodic training structure is now given by (30) with K replaced by K^{*}.

IV. Optimal Training Subblock Signal

In the previous section, we designed the optimal periodic training *structure* for a fixed block length and a fixed *received* training signal energy. Although the channel dispersion effect is included in the design, the channel fading effect is excluded from the design due to the condition of the *fixed received* training signal energy. The snapshot CRB, which is a function of the (snapshot) received training signal energy, was used in the training *structure* design. In this section, we adopt minimizing the *average* CRB as the training design criterion in order to include the channel fading effect. The average CRB is the average of the snapshot CRB over the received signal energy statistics (or equivalently, over the fading channel statistics). Note that the average CRB corresponds to the extended Miller and Chang bound (EMCB) proposed in [19], where it was reported that EMCB can sometimes be tighter than the CRB.[5] We will investigate the effect of the channel fading on the average CRB, and we will find an optimal training signal within a training subblock so that the average CRB is minimized. We consider a frequency-selective multipath Rayleigh fading channel. The problem can be formulated as follows: "For a fixed *transmitted* energy $\mathcal{E}_{\mathrm{Tx}}$ of a periodic training signal composed of several identical training subblocks with the subblock energy $\mathcal{E}_{\mathrm{T1}}$, what is the best training subblock signal that minimizes the average CRB in a frequency-selective multipath Rayleigh fading channel?"

[5]EMCB is tighter than the hybrid CRB, modified CRB, MCB, and asymptotic CRB, but its application is restricted to locally unbiased estimators, i.e., those estimators that are unbiased for all values of the nuisance parameters [19]. We use a best linear unbiased estimation (BLUE) method which is locally unbiased regardless of the channel conditions at the SNRs of practical interest, given that the frequency offset is within the estimation range of the method.

The multipath channel fading causes fluctuation of the received training signal energy (although the long-term average or the expected value of the received training signal energy is fixed), which, in turn, affects the average CRB. To minimize the average CRB, the training signal should be designed such that the received training signal energy fluctuation is minimized. This fact will be proven later. In the context of a periodic training signal, only one training subblock needs to be considered in the training signal design.

Define the following:

$$Z=\sum_{k=0}^{L-1}|x(k)|^2 \tag{38}$$

which represents the (snapshot) energy of one received training subblock. Then for a given periodic training structure, we know from (22) that

$$\mathrm{CRB}_{|\mathrm{SNR}_{ss}}=\frac{\alpha}{Z}=\mathrm{CRB}_{|Z} \tag{39}$$

where

$$\alpha=\frac{\sigma_n^2}{8\pi^2L^2\left(\sum_{k=0}^{U-1}i_k^2-\frac{\left(\sum_{n=0}^{U-1}i_n\right)^2}{U}\right)}. \tag{40}$$

In a multipath Rayleigh fading channel, $\{x(k):k=0,\ldots,L-1\}$ are complex Gaussian random variables. The probability density function (pdf) of Z can be given by

$$p_Z(z)=\sum_{l=1}^{m}\sum_{k=1}^{\kappa_l}A_{lk}\frac{1}{\lambda_l}f_{2k}(z/\lambda_l) \tag{41}$$

where $f_n(\ )$ is a Chi-square pdf with n degrees of freedom (DOFs), m is the number of nonzero distinct eigenvalues $\{\lambda_l\}$ of $\boldsymbol{C}_{\boldsymbol{h}}^{1/2}\boldsymbol{S}_o^H\boldsymbol{S}_o\boldsymbol{C}_{\boldsymbol{h}}^{1/2}$, and κ_l is the multiplicity of λ_l (see Appendix D for details on the parameters and derivation). The average CRB is defined as

$$\mathrm{CRB}=\int_{\beta}^{\infty}\mathrm{CRB}_{|Z}p_Z(z)dz=\int_{\beta}^{\infty}\frac{\alpha}{z}p_Z(z)dz \tag{42}$$

where β is typically zero. Note that if one wants a more practical performance limit, then β can be set to a positive value relating to the receiver sync detection threshold. After some manipulation, we obtain

$$\mathrm{CRB}=\sum_{l=1}^{m}\left\{-\frac{A_{l1}}{\lambda_l}Ei\left(\frac{-\beta}{\lambda_l}\right)+\sum_{k=2}^{\kappa_l}\frac{A_{lk}}{(k-1)\lambda_l}\left(1-F_{2k-2}(\beta)\right)\right\} \tag{43}$$

where $Ei(x)$ is the exponential integral function and $F_{2k}(x)$ is the cumulative distribution function of a Chi-square random variable with $2k$ DOFs, respectively defined by

$$Ei(x) = -\int_{-x}^{\infty} \frac{e^{-u}}{u} du \tag{44}$$

$$F_{2k}(x) = \int_{0}^{x} f_{2k}(u) du. \tag{45}$$

Then the best training signal is determined by the parameters $[m, \{\kappa_l\}, \{\lambda_l\}]^*$ that minimize the average CRB as

$$\begin{aligned}&[m, \{\kappa_l\}, \{\lambda_l\}]^* \\ &\quad = \underset{m,\{\kappa_l\},\{\lambda_l\}}{\arg\min}\ \text{CRB} \\ &\qquad \text{constraints}: \sum_{l=1}^{m} \kappa_l \le L;\ \text{trace}\left\{\boldsymbol{S}_o^H \boldsymbol{S}\right\} = L\mathcal{E}_{\text{T1}}; \\ &\qquad \lambda_l > 0;\ 1 \le m \le L;\ \kappa_l \in \{1, 2, \ldots, L\}.\end{aligned} \tag{46}$$

However, solving the above problem appears to be intractable. Hence, in the following, we resort to a simpler approach using an approximate average CRB. If $\{x(k) : k = 0, \ldots, L-1\}$ are circularly symmetric, independent and identically distributed (i.i.d.), then Z is a summation of $2L$ i.i.d. Chi-square random variables and is a Chi-square random variable with $2L$ DOFs. If $\{x(k) : k = 0, \ldots, L-1\}$ are correlated, then Z is a summation of correlated Chi-square random variables which can be well approximated by a Gamma random variable with the parameter n_g (see [20] and references therein). The Gamma distribution is a generalization of the Chi-square distribution in that n_g is an integer (representing the DOFs) in the latter, but can be any positive real number in the former. In our case, $n_g \ge 2$. The mean and variance of Z are given by $E[Z] = n_g\sigma_g^2$ and $\sigma_Z^2 = 2n_g\sigma_g^4$, where the values of n_g and σ_g^2 depend on the training signal, but $E[Z] = n_g\sigma_g^2 = \sum_{k=0}^{L-1} |s_k|^2 = \mathcal{E}_{\text{T1}}$ is a constant representing the transmitted energy of one training subblock. Then the approximate average CRB in the multipath Rayleigh fading channel can be given by

$$\begin{aligned}\text{CRB} &\simeq \int_0^{\infty} \frac{\alpha}{z} \frac{1}{\sigma_g^{n_g} 2^{n_g/2}\Gamma(n_g/2)} z^{\frac{n_g}{2}-1} e^{-\frac{z}{2\sigma_g^2}} dz \\ &= \frac{\alpha\Gamma\left(\frac{n_g}{2}-1\right)}{2\sigma_g^2\Gamma\left(\frac{n_g}{2}\right)} \int_0^{\infty} \frac{1}{\sigma_g^l 2^{l/2}\Gamma(l/2)} z^{\frac{l}{2}-1} e^{-\frac{z}{2\sigma_g^2}} dz \\ &= \frac{\alpha\Gamma\left(\frac{n_g}{2}-1\right)}{2\sigma_g^2\Gamma\left(\frac{n_g}{2}\right)} = \frac{\alpha}{n_g\sigma_g^2 - 2\sigma_g^2} \\ &= \frac{\alpha}{E[Z] - \sigma_Z^2/E[Z]} = \frac{\alpha}{\mathcal{E}_{\text{T1}} - \sigma_Z^2/\mathcal{E}_{\text{T1}}}\end{aligned} \tag{47}$$

where $l = n_g - 2$ and we have used $\Gamma(x+1) = x\Gamma(x)$.

Equation (47) indicates that the larger fluctuation of the received training signal energy (larger σ_Z^2) causes the larger average CRB in the multipath Rayleigh fading channel. In other words, the training signal which gives the minimum fluctuation of the received training signal energy is the optimal training signal.

After some calculation, the variance of Z is given by (see Appendix E for details)

$$\begin{aligned}\sigma_Z^2 = &\sum_{k=0}^{L-1} \left(\boldsymbol{C}_{\boldsymbol{S}}(k,k)\right)^2 2\sigma_{h_k}^4 \\ &+ \sum_{m=0}^{L-1} \sum_{n=0, n\ne m}^{L-1} \boldsymbol{C}_{\boldsymbol{S}}(m,n)\boldsymbol{C}_{\boldsymbol{S}}(n,m)\sigma_{h_m}^2\sigma_{h_n}^2 \\ &+ \sum_{i=0}^{L-1} \sum_{l=0, l\ne i}^{L-1} \boldsymbol{C}_{\boldsymbol{S}}(i,i)\boldsymbol{C}_{\boldsymbol{S}}(l,l)\sigma_{h_i}^2\sigma_{h_l}^2 - \mathcal{E}_{\text{T1}}^2\end{aligned} \tag{48}$$

where $\boldsymbol{C}_{\boldsymbol{S}}(m,n)$ is the (m,n)th element of $\boldsymbol{C}_{\boldsymbol{S}} = \boldsymbol{S}_o^H\boldsymbol{S}_o$. Note that $\boldsymbol{C}_{\boldsymbol{S}}$ represents the periodic autocorrelation matrix of the training signal $\{s_0, s_1, \ldots, s_{L-1}\}$ and $\boldsymbol{C}_{\boldsymbol{S}}(n,n) = \sum_{k=0}^{L-1} |s_k|^2$ for all n and $\boldsymbol{C}_{\boldsymbol{S}}(m,n) = \boldsymbol{C}_{\boldsymbol{S}}^*(n,m)$, hence, all terms (except $-\mathcal{E}_{\text{T1}}^2$) in (48) have real nonnegative values. From (48), it is clear that σ_Z^2 is minimized when $\boldsymbol{C}_{\boldsymbol{S}}(l,k) = 0$ for $l \ne k$. In other words, the training signal which possesses zero periodic autocorrelation for any nonzero correlation lag (usually referred to as a ZAC signal) minimizes the received training signal energy fluctuation in the multipath Rayleigh fading channel. Combining this result with that of (47) implies that the ZAC training signals are optimal for frequency offset estimation in multipath Rayleigh fading channels.

For ZAC training signals, (48) becomes

$$\begin{aligned}\sigma_Z^2 &= \mathcal{E}_{\text{T1}}^2 \left[\sum_{k=0}^{L-1} 2\sigma_{h_k}^4 + \sum_{i=0}^{L-1} \sum_{l=0, l\ne i}^{L-1} \sigma_{h_i}^2\sigma_{h_l}^2 - 1\right] \\ &= \mathcal{E}_{\text{T1}}^2 \sum_{k=0}^{L-1} \sigma_{h_k}^4\end{aligned} \tag{49}$$

and the corresponding average CRB is given by

$$\text{CRB} = \frac{\alpha}{\mathcal{E}_{\text{T1}}\left(1 - \sum_{k=0}^{L-1} \sigma_{h_k}^4\right)}. \tag{50}$$

From (49) and (50), we observe that for ZAC training signals, σ_Z^2, and hence, CRB just depend on the transmitted energy and the channel power delay profile, not on the detailed characteristics of the training signals (except the ZAC property). Note that the variance of Z can also be obtained from (D-2) as

$$\sigma_Z^2 = \sum_{k=0}^{L-1} \gamma_k^2 \tag{51}$$

where $\{\gamma_k\}$ are eigenvalues of $\boldsymbol{C}_{\boldsymbol{h}}^{1/2}\boldsymbol{S}_o^H\boldsymbol{S}_o\boldsymbol{C}_{\boldsymbol{h}}^{1/2}$ [see (D-1) and (D-4)].

For single-carrier systems, constant amplitude ZAC (CAZAC) sequences will be a preferred choice for optimal training subblock signals. The construction of CAZAC

sequences can easily be found in the literature (e.g., [6] and references therein). Since $\boldsymbol{S}_o$ is an $L \times L$ circulant matrix, it can be diagonalized by the discrete Fourier transform (DFT) matrix as

$$\boldsymbol{S}_o = \boldsymbol{F}_L^H \boldsymbol{\Sigma} \boldsymbol{F}_L \tag{52}$$

where $\boldsymbol{\Sigma}$ is a diagonal matrix containing eigenvalues of $\boldsymbol{S}_o$ and $\boldsymbol{F}_L$ is the L-point unitary DFT matrix. The ZAC property of the time-domain signal can be expressed as

$$\boldsymbol{S}_o^H \boldsymbol{S}_o = \mathcal{E}_{\mathrm{T1}} \boldsymbol{I}_L. \tag{53}$$

Substituting (52) into (53) gives

$$\boldsymbol{\Sigma}^H \boldsymbol{\Sigma} = \mathcal{E}_{\mathrm{T1}} \boldsymbol{I}_L. \tag{54}$$

For orthogonal frequency-division multiplexing (OFDM) systems, the ZAC property of a time-domain subblock signal can be related to its frequency-domain counterpart as follows. Let $\{c_0, c_1, \ldots, c_{L-1}\}$ be the L-point (unitary) DFT of the time-domain training signal $\{s_0, s_1, \ldots, s_{L-1}\}$. Then it can be proven that $\boldsymbol{\Sigma} = \sqrt{L}\mathrm{diag}\{c_0, c_1, \ldots, c_{L-1}\}$. Together with (54), we can conclude that the ZAC property of the optimal time-domain subblock signal $\{s_0, s_1, \ldots, s_{L-1}\}$ translates to the frequency-domain property that $|c_k|^2$ is constant for all k. An optimal ZAC time-domain subblock signal for OFDM systems can then be generated by the L-point inverse (I)DFT of $\{c_0, c_1, \ldots, c_{L-1}\}$ where $\{c_k\}$ have equal energy. The phase of $\{c_k\}$ can be designed to achieve low peak-to-average energy ratio (PAR) of the training signal (to avoid nonlinear distortion at the power amplifier). For example, a CAZAC sequence $\{c_k\}$ will give a very low PAR, since the low correlation property of frequency-domain subcarrier symbols translates into low PAR of the time-domain signal.

It is worth noting that the ZAC property of the proposed training signal also satisfies $\boldsymbol{S}^H \boldsymbol{S} = \mathcal{E}_{\mathrm{Tx}} \boldsymbol{I}$, the optimality criterion of the least-squares (also maximum likelihood, in this case) channel estimation (see [11]). Hence, the proposed subblock signal is optimal for estimation of both frequency offset and channel. On the other hand, the proposed training structure sacrifices energy of one subblock if compared with the conventional training structure with consecutive identical subblocks. Although the proposed structure is the best for frequency offset estimation, it may not be the best for the channel estimation due to the energy loss. As discussed before, the best training structure and signal for estimation of both frequency offset and channel is a delicate open problem and is beyond the scope of this paper. However, in the following section, we evaluate the estimation performance and data error performance for the systems using the same training signal in estimation of both frequency offset and channel, and show that the proposed training structure with a ZAC signal gives a better BER result than the conventional training structure with a ZAC signal.

V. Simulation Results and Discussions

A. Simulation Parameters

The channel is assumed to have an exponential power delay profile with a 3 dB per tap decaying factor. We assume that the channel gains remain unchanged over a block of length $40L$ samples (i.e., $P = 39$) where the channel length L is 16 samples, unless stated otherwise.

We first evaluate various training structures which have the same total transmitted training energy $\mathcal{E}_{\mathrm{Tx}}$ and are composed of identical training subblocks of length L samples each. The conventional structure contains $(K_c + 1)$ consecutive identical subblocks. The first subblock is discarded and the remaining K_c subblocks are used in the estimation. The proposed structure is the one in which the first and the last $K + 1$ subblocks are training signals and $2K$ training subblocks are used in the estimation. Note that the conventional structure with $K_c = 19$ corresponds to the training signal structure proposed in [21] as an improvement over the IEEE 802.11a training structure. Next, we evaluate several training subblock signals, namely, a CAZAC sequence, a bipolar m-sequence (one bit appended to have an even length $L = 16$ and a full-rank training matrix), the IEEE 802.11a short preamble, the all-one sequence, an alternating ± 1 sequence, and a full-rank, non-ZAC signal (the L eigenvalues of the corresponding training matrix are proportional to k^2 where $k = 1, \ldots, L$).

The frequency offset estimation method of [22] (Method-B) is used in the evaluation of the training structures. For very low SNR, the initial frequency offset compensation of [22] is done by the method of [21] as it gives a slightly better performance than the initial frequency offset estimator of [22]. The results are presented for different values of $\gamma = \mathcal{E}_{\mathrm{Tx}}/(160\sigma_n^2)$. For the proposed structure with $K = 4$ and the conventional structure with $K_c = 9$, γ equals SNR. We set the normalized frequency offset $v = 0.4/64$ (For an OFDM system with 64 subcarriers, this frequency offset corresponds to 0.4 times the subcarrier spacing.) The simulation results are obtained by 10^4 independent channel realizations unless stated otherwise. For BER evaluation, we consider an OFDM system (for the sake of low-complexity equalization) with 64 subcarriers, and each packet contains a training signal and 20 OFDM data symbols with binary phase-shift keying (BPSK) modulation. For coded BER performance, we use the industrial standard rate 1/2, constraint length 7, convolutional code with the generator matrix $(133_8, 171_8)$, and two-stage interleaver from [26]. We applied the soft-decision Viterbi decoding with branch metric $\sum_{i=0}^{1} \|Y_{2k+i} - \hat{H}_{2k+i} X_i\|^2$ where Y_{2k+i} and $\hat{H}_{2k+i}$ are the received subcarrier symbol and corresponding channel gain estimate after the deinterleaver, and $\{X_i : i = 0, 1\}$ represent the two output symbols on the considered trellis branch.

B. Discussion on the Training Structure

Fig. 2 presents the advantage of the proposed structure over the conventional structure in terms of the snapshot CRB. Under the same training signal energy constraint, a larger K_c, which corresponds to more system resource (time duration), gives a better result for the conventional structure. When both structures

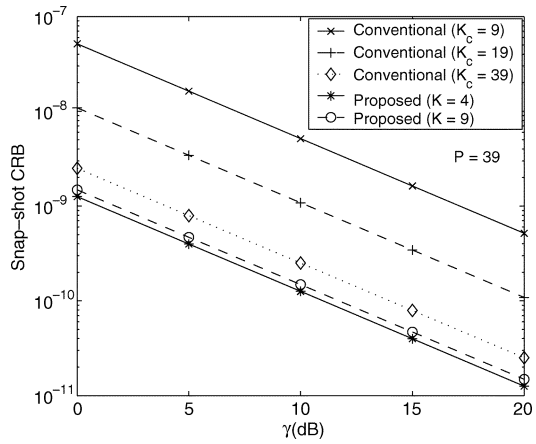


Fig. 2. Snapshot CRB of the normalized frequency offset estimation with different periodic training structures.

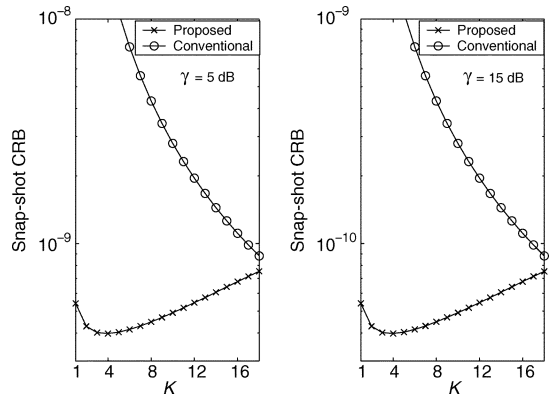


Fig. 3. Snapshot CRB for different periodic training structures as a function of K (for the conventional structure, $K_c = 2K + 1$).

use the same system (time) resource (i.e. when $K_c = 2K + 1$), the advantage of the proposed structure is quite significant.

In Fig. 3, the snapshot CRBs for both structures using the same system resource are plotted as a function of K where $K_c = 2K + 1$. The advantage of the proposed structure is more significant for smaller K. For the conventional structure, a larger K_c gives better performance which improves significantly as K_c increases. For the proposed structure, there is an optimal K value (in this case, it is four), and any decrease or increase from the optimal value results in a slight performance degradation. However, the performance differences for different K values in the proposed structure are not as significant as they are in the conventional structure. Optimal K values for several P values are plotted in Fig. 4.

Fig. 5 shows the simulation results of the normalized frequency offset estimation mean-square errors (MSEs) in the considered frequency-selective fading channel for both structures (with CAZAC subblock signal). At high SNR, the proposed structure achieves a better MSE result than the conventional structure and the optimal value $K = 4$ gives a slightly better performance than $K = 9$. But at low SNR, $K = 9$ yields a better result. At low SNR, the proposed structure with a smaller K

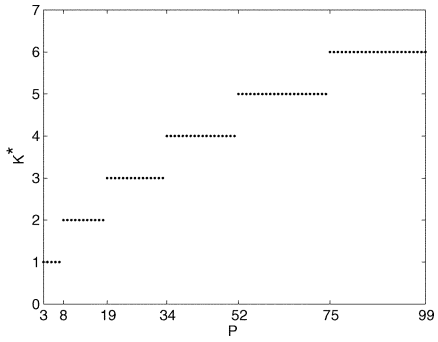


Fig. 4. Optimal value of K versus P for the optimal training structure containing $(2K + 2)$ subblocks within the block length of $P + 1$ subblocks. (The plot is continuous from the right, e.g., at $P = 34$, $K^* = 4$.)

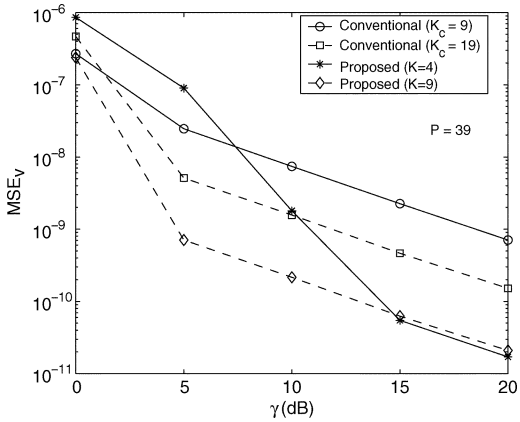


Fig. 5. Normalized frequency offset estimation MSE for different periodic training structures.

gives a larger MSE than the conventional structure.[6] This can be explained by the frequency offset estimation likelihood metrics of the different training structures. The proposed structure has a sharper mainlobe of the likelihood metric trajectory (which gives better estimation accuracy) but it also has larger sidelobes which result in a large estimation error when the received signal is in deep fade. Since the differences in snapshot CRB for the proposed structure with different K values are not significant, the large sidelobe effect of the proposed structure suggests that for a robust performance, a larger K value than the optimal one should be used in practice. On the other hand, a larger K value requires more system resource (time duration).

A robust K design is given in Appendix G (similar to the design in [23] except with different metric) where K is chosen such that the sidelobe peak of the estimation metric is less than 0.75 of the mainlobe peak. $K = 9$ satisfies this design criterion. It should be noted that the sidelobe effect (or outlier) would be smaller for practical systems since the packet in deep fade would not be detected and the corresponding outlier present in the simulation would not be observed in practice. Another practical design issue is the value of P which should be chosen such that the channel is time-invariant over the $P + 1$ subblocks.

The uncoded BER simulation results are shown in Fig. 6 for both structures (SNR $= E_b/N_0$ in this case). The proposed structure gives a better BER peroformance. It should be noted

[6]A similar effect occurs in the training design of [14] for AWGN channel.

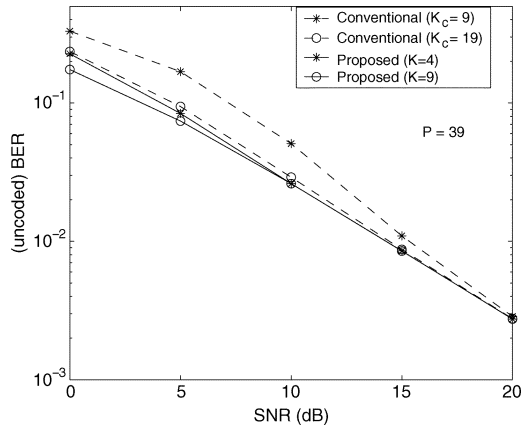


Fig. 6. Uncoded BER performance obtained with different training structures.

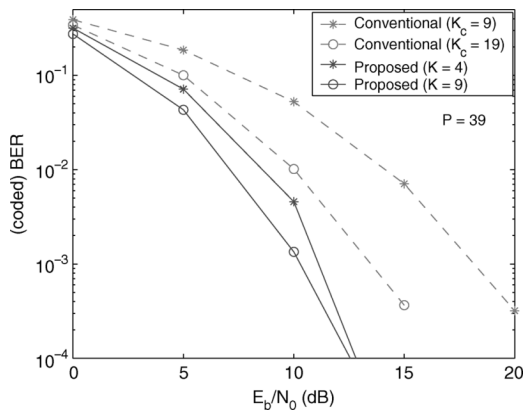


Fig. 7. Coded BER performance obtained with different training structures (rate-1/2 convolutional code).

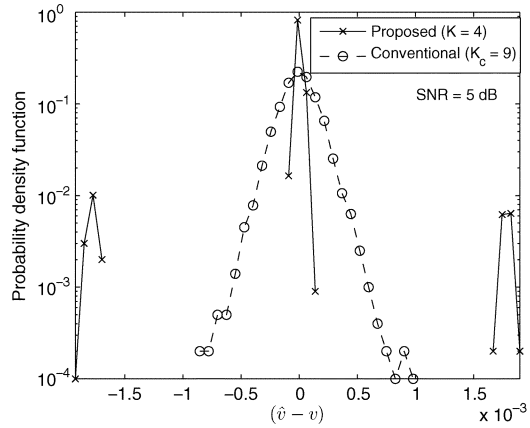


Fig. 8. PDF (obtained by simulation) of the normalized frequency offset estimation error for different training structures.

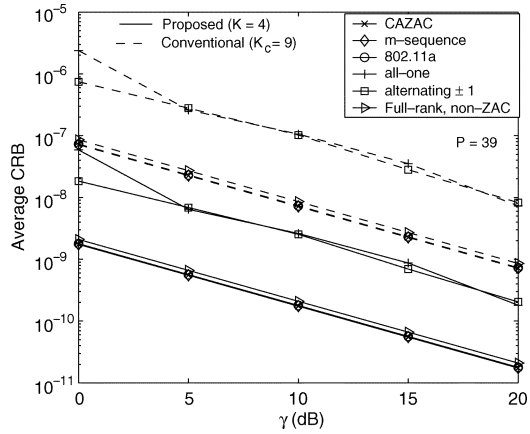


Fig. 9. Average CRB of the normalized frequency offset estimation for different training structures with different subblock signals in a multipath Rayleigh fading channel.

that even though the proposed structure with $K = 4$ gives a larger frequency offset estimation MSE than the conventional structure, it achieves a better BER. The reason can be explained by the pdf of the frequency offset estimation error shown in Fig. 8 (obtained by simulation). The proposed structure with $K = 4$ gives a more accurate estimate most of the time, but occasional outliers due to its large likelihood metric sidelobes cause large errors resulting in a larger MSE than the conventional structure. The relationship between BER and frequency offset estimation MSE is nonlinear and the effect of occasional outliers is less significant on BER than on MSE. This results in a better BER performance for the proposed structure.

The coded BER results are plotted in Fig. 7 where E_b/N_0 (of the information bit) is 3 dB larger than the SNR due to rate-1/2 coding with BPSK modulation. The proposed structure brings in a much greater BER performance gain in coded systems than it does in uncoded systems over the conventional structure. In particular, at high SNR, although the uncoded BERs of the two structures are quite close, the performance gap for the coded BER between the two structures is quite significant. This fact illustrates the significant impact of frequency offset estimation accuracy on the soft-decision Viterbi decoding.

C. Discussion on the Training Subblock Signal

Fig. 9 presents average CRBs corresponding to several subblock signals for the proposed structure with $K = 4$ and the conventional structure with $K_c = 9$. The results are obtained from 10^7 simulation runs. The proposed subblock signal (ZAC signal or a full-rank training matrix with constant amplitude eigenvalues) performs the best. Both the m-sequence and the 802.11a short preamble have correlation properties very close to the ZAC property, hence their average CRBs are almost the same as the average CRB of the CAZAC sequence. The average CRB for the full-rank, non-ZAC sequence is also close to that of the CAZAC sequence. The other two subblock signals have a large autocorrelation and hence, their average CRBs are significantly larger than the average CRB of the CAZAC sequence. In Fig. 10, the normalized frequency offset estimation MSEs corresponding to several training subblock signals are presented. The MSE performance follows the same trend as the average CRB.

The LS channel estimation MSEs (simulation results) obtained with several training subblock signals are plotted in Fig. 11 for both training structures. The theoretical channel estimation MSE is given in Appendix F. Note that the signal dimension within which the LS method operates equals the

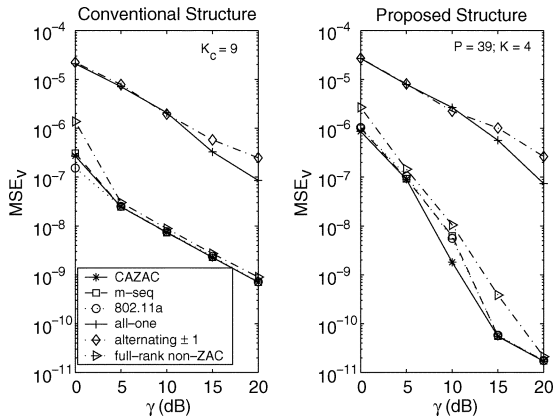


Fig. 10. Normalized frequency offset estimation MSE of different training subblock signals in a multipath Rayleigh fading channel.

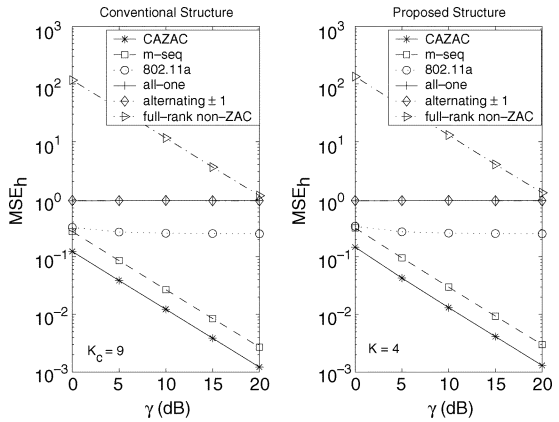


Fig. 11. Channel estimation MSE of different training subblock signals in a multipath Rayleigh fading channel.

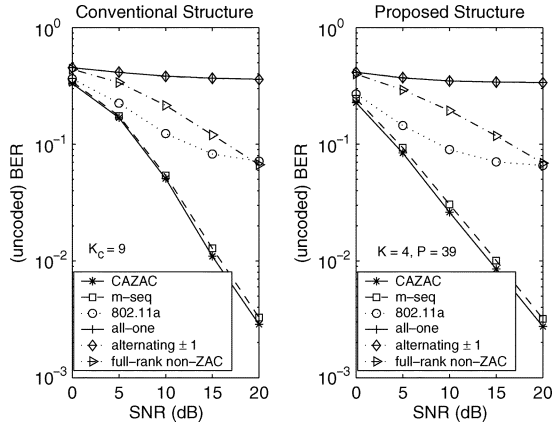


Fig. 12. Uncoded BER performance of different training subblock signals in a multipath Rayleigh fading channel.

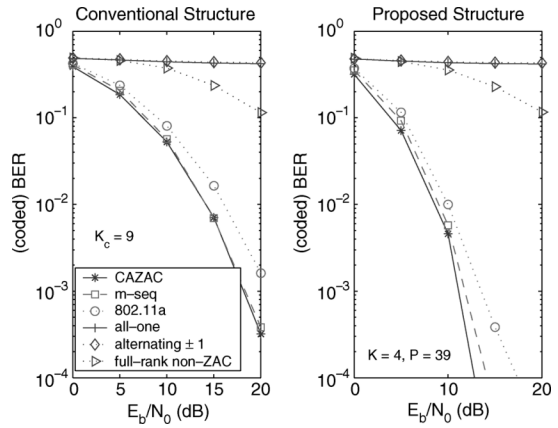


Fig. 13. Coded BER performance of different training subblock signals in a multipath Rayleigh fading channel (rate-1/2 convolutional code).

rank of the training matrix. Since the 802.11a short preamble, the all-one sequence, and the alternating ± 1 sequence are rank-deficient, their MSEs are dominated by the errors due to the missing channel coefficients from the missing dimensions, hence causing MSE floors [see also the first term of (F-7)]. Since the ZAC property is the optimality criterion for the LS method, the CAZAC sequence gives the best MSE performance. Although the full-rank, non-ZAC signal has a full rank, thus no MSE floor, the eigenvalues of the corresponding training matrix is far away from the optimal values (constant amplitude eigenvalues) resulting in a large MSE [see also the second term of (F-7)].

The uncoded BER results corresponding to several training subblock signals are shown in Fig. 12 for both training structures. The proposed ZAC subblock signal gives the best BER. Since the autocorrelation property of the m-sequence used is very close to the ZAC property, the corresponding BER is close to that of CAZAC sequence. A larger BER degradation is observed for a subblock signal with autocorrelation property less similar to the ZAC property.

The coded BER results for several training subblock signals are presented in Fig. 13. Similar to uncoded BER results, the CAZAC sequence gives the best performance for both structures. A difference from the uncoded BER results is that the IEEE 802.11a preamble gives a substantial improvement in coded systems (due to its closeness to ZAC property, the optimal subblock signal criterion for frequency offset estimation). Another observation is that the advantage of the optimal training subblock signal is slightly greater with the proposed structure than with the conventional structure.

D. *Effect of Channel Order Mismatch*

Next, we investigate the effect of mismatch in the number of actual channel taps L and the design value for L used in the estimation of frequency offset and channel. We use a fixed design value of 16 with various L values. The normalized frequency offset estimation MSE, the channel estimation MSE, and the BER results are plotted, respectively, in Figs. 14–16 for both training structures using the CAZAC subblock signal. No significant performance loss is observed for the mismatched value of L.

From Figs. 14–16, we can also compare the proposed structure and the conventional structure both using CAZAC subblock signal and the same system resource (the same number of training subblocks transmitted). The proposed structure has a significantly better frequency offset estimation MSE (except at low SNR) but a slightly larger channel estimation MSE (for larger K, the difference is not noticeable) due to the training energy loss associated with the proposed structure. In terms of

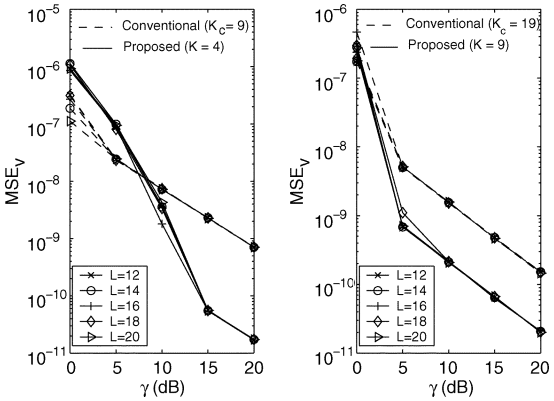


Fig. 14. Normalized frequency offset estimation MSE of several training signals with mismatched L values in a multipath Rayleigh fading channel (The design value for L is 16.)

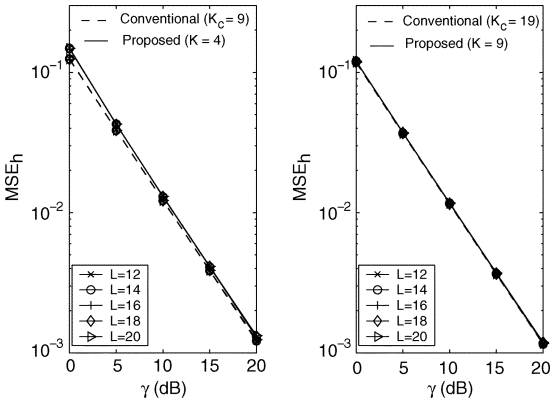


Fig. 15. Channel estimation MSE of different training subblock signals with mismatched L values in a multipath Rayleigh fading channel (the design value for L is 16.)

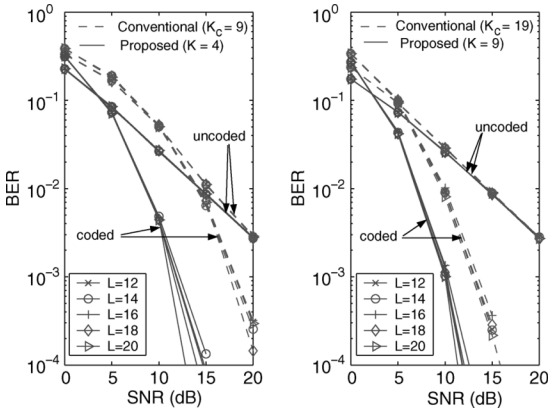


Fig. 16. Uncoded BER performance of different training signal structures with mismatched L values in a multipath Rayleigh fading channel (the design value for L is 16.)

BER performance, the proposed structure gives a better result which is more significant for the coded BER.

E. Discussion on Closed Form (Approximate) Average CRB

Another contribution in this paper is a simple closed form (approximate) average CRB in multipath Rayleigh fading channels. For ZAC (optimal) signals, our closed form average CRB given in (50) is the exact expression. For non-ZAC signals, it is an approximate expression [easily computed from (47) and (51)] and

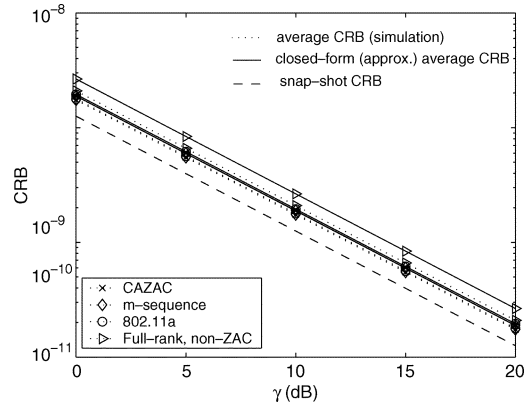


Fig. 17. Comparison of the closed form (approximate) average CRB, the average CRB obtained by simulation, and the snapshot CRB for the proposed training structure with $K = 4$.

is not applicable for the rank-one training matrix $\boldsymbol{S}_o$. We evaluated the average CRB in the multipath Rayleigh fading channel by Monte Carlo simulation (i.e., average of the snapshot CRBs over independent channel realizations) and compared with our closed-form CRB. The results are plotted in Fig. 17. The commonly used snapshot CRB is also included in the comparison. We observe that the closed-form CRBs match the simulation results better than the snapshot CRB does. The snapshot CRB serves as a lower bound for the average CRB. For non-ZAC signals, the approximate closed-form results are observed to be slightly larger than the simulation results, but they give a better approximation to the average CRB than the snapshot CRB.

VI. Conclusions

We have presented a periodic training signal design for frequency offset estimation in a multipath fading channel in terms of the optimal periodic training structure (the optimal locations of identical training signal subblocks) and the optimal training subblock signal. The proposed training structure contains $K + 1$ identical training subblocks each having length L samples (the channel length) located at the beginning and the remaining $K + 1$ identical training subblocks located at the end of the block interval during which the channel remains unchanged. The optimal value of K can be easily computed and gives the best results at high SNR. We also presented a robust design for K value to reduce the occurrence of outliers at low SNR. The optimal training subblock signal has a ZAC property which results in the minimum fluctuation of received energy in frequency-selective Rayleigh fading channels which in turn translates into the minimum average CRB. A simple closed-form average CRB (exact for ZAC signals and a good approximate for non-ZAC signals) in multipath Rayleigh fading channels is also presented which gives a better bound than the common snapshot CRB. Both the training structure and the subblock signal characteristics are important since large deviations from the optimal structure or optimal subblock signal characteristics result in large performance degradation. In light of the existing training structures and subblock signals, the proposed training structure will have a greater impact on the estimation improvement since the existing training subblock signals already have ZAC or close-to-ZAC properties. Finally, we observed that the

advantages of the proposed training structure and the proposed training subblock signal over conventional ones are greater in coded systems than they are in uncoded systems.

APPENDIX A

In this appendix, we derive the snapshot CRB of frequency offset estimation using an arbitrary periodic received training structure shown in Fig. 1. Within the P subblocks, U subblocks with block indexes $\{i_0, i_1, \ldots, i_{U-1}\}$ are used in the estimation and the observation vector $\boldsymbol{y}$ is given in (17). The likelihood function, given the parameters $(\boldsymbol{h}, v)$, takes the form

$$\boldsymbol{p}(\boldsymbol{y}; \tilde{\boldsymbol{h}}, \tilde{v}) = \frac{1}{(\pi\sigma_n^2)^{UL}} \cdot \exp\left(\frac{-1}{\sigma_n^2}\left\|\boldsymbol{y} - \boldsymbol{\Gamma}(\tilde{v})\boldsymbol{C}\tilde{\boldsymbol{h}}\right\|^2\right) \quad \text{(A-1)}$$

where $\|\cdot\|^2$ is the vector norm-square, $\tilde{\boldsymbol{h}}$ and $\tilde{v}$ are trial values of $\boldsymbol{h}$ and v. Following the same approach as in [18], we obtain the snapshot CRB for the estimation of v as

$$\text{CRB}|_{\boldsymbol{h}} = \frac{\sigma_n^2}{2\boldsymbol{z}^H(\boldsymbol{I}_{UL} - \boldsymbol{D})\boldsymbol{z}} \quad \text{(A-2)}$$

where

$$\boldsymbol{z} = 2\pi\boldsymbol{M}\boldsymbol{C}\boldsymbol{h} \quad \text{(A-3)}$$

$$\boldsymbol{M} = \boldsymbol{Q}^T\boldsymbol{\Lambda}\boldsymbol{Q} \quad \text{(A-4)}$$

$$\boldsymbol{D} = \boldsymbol{C}(\boldsymbol{C}^H\boldsymbol{C})^{-1}\boldsymbol{C}^H. \quad \text{(A-5)}$$

After some manipulation, we have

$$\boldsymbol{M} = \text{diag}\{i_0 L\boldsymbol{I}_L + \boldsymbol{M}_L, i_1 L\boldsymbol{I}_L + \boldsymbol{M}_L, \ldots, i_{U-1}L\boldsymbol{I}_L + \boldsymbol{M}_L\} \quad \text{(A-6)}$$

$$\boldsymbol{M}_L = \text{diag}\{0, 1, \ldots, L-1\}$$

$$\boldsymbol{D} = \frac{1}{U}\boldsymbol{1}_{U\times U} \otimes \bar{\boldsymbol{D}}_1 \quad \text{(A-7)}$$

$$\bar{\boldsymbol{D}}_1 = \boldsymbol{S}_o\left(\boldsymbol{S}_o^H\boldsymbol{S}_o\right)^{-1}\boldsymbol{S}_o^H = \bar{\boldsymbol{F}}_L^H\boldsymbol{\Psi}\bar{\boldsymbol{F}}_L \quad \text{(A-8)}$$

$$\boldsymbol{z} = 2\pi\left[(\boldsymbol{M}_L\boldsymbol{S}_o\boldsymbol{h} + i_0 L\boldsymbol{S}_o\boldsymbol{h})^T, (\boldsymbol{M}_L\boldsymbol{S}_o\boldsymbol{h} + i_1 L\boldsymbol{S}_o\boldsymbol{h})^T, \ldots, (\boldsymbol{M}_L\boldsymbol{S}_o\boldsymbol{h} + i_{U-1}L\boldsymbol{S}_o\boldsymbol{h})^T\right]^T \quad \text{(A-9)}$$

where $\boldsymbol{1}_{U\times U}$ is the $U \times U$ all-one matrix and $\boldsymbol{\Psi}$ is a diagonal matrix with ones at non-zero eigenvalues of $\boldsymbol{S}_o$ and zeros otherwise (i.e., $\boldsymbol{\Psi}$ is obtained by replacing nonzero elements of $\boldsymbol{\Sigma}$ from (52) with ones). Note that in (A-8), a pseudoinverse is applied if $\boldsymbol{S}_o$ is not of full rank. Then the terms from (A-2) are given by

$$\boldsymbol{z}^H\boldsymbol{z} = 4\pi^2\left(U\boldsymbol{h}^H\boldsymbol{S}_o^H\boldsymbol{M}_L^2\boldsymbol{S}_o\boldsymbol{h} + L^2\sum_{k=0}^{U-1} i_k^2\boldsymbol{h}^H\boldsymbol{S}_o^H\boldsymbol{S}_o\boldsymbol{h} + 2L\sum_{k=0}^{U-1} i_k\boldsymbol{h}^H\boldsymbol{S}_o^H\boldsymbol{M}_L\boldsymbol{S}_o\boldsymbol{h}\right) \quad \text{(A-10)}$$

$$\boldsymbol{z}^H\boldsymbol{D}\boldsymbol{z} = 4\pi^2\left\{U\boldsymbol{h}^H\boldsymbol{S}_o^H\boldsymbol{M}_L^H\boldsymbol{D}_1\boldsymbol{M}_L\boldsymbol{S}_o\boldsymbol{h} + \frac{L^2}{U}\left(\sum_{k=0}^{U-1} i_k\right)^2\boldsymbol{h}^H\boldsymbol{S}_o^H\boldsymbol{D}_1\boldsymbol{S}_o\boldsymbol{h} + 2L\sum_{k=0}^{U-1} i_k\boldsymbol{h}^H\boldsymbol{S}_o^H\boldsymbol{M}_L\boldsymbol{D}_1\boldsymbol{S}_o\boldsymbol{h}\right\} \quad \text{(A-11)}$$

$$\boldsymbol{z}^H(\boldsymbol{I}_{UL} - \boldsymbol{D})\boldsymbol{z} = 4\pi^2\boldsymbol{h}^H\boldsymbol{F}_L^H \times\left\{U\boldsymbol{\Sigma}^H\bar{\boldsymbol{F}}_L\boldsymbol{M}_L\bar{\boldsymbol{F}}_L^H(\boldsymbol{I}_L - \boldsymbol{\Psi})\bar{\boldsymbol{F}}_L\boldsymbol{M}_L\bar{\boldsymbol{F}}_L^H\boldsymbol{\Sigma} + L^2\left[\sum_{k=0}^{U-1} i_k^2 - \frac{\left(\sum_{n=0}^{U-1} i_n\right)^2}{U}\right]\boldsymbol{\Sigma}^H\boldsymbol{\Sigma}\right\}\boldsymbol{F}_L\boldsymbol{h}. \quad \text{(A-12)}$$

Finally, substituting (A-12) into (A-2) gives the snapshot CRB. Note that this CRB is derived for joint estimation of frequency offset and channel. The identifiability of an L-tap channel requires a full-rank $\boldsymbol{S}_o$. A rank-deficient $\boldsymbol{S}_o$ results in a performance degradation as can be seen from the simulation results or theoretical MSE derived in Appendix E. Hence, we consider a full-rank $\boldsymbol{S}_o$ in our design. Then $\boldsymbol{\Psi} = \boldsymbol{I}_L$ and the snapshot CRB is given by

$$\text{CRB}|_{\boldsymbol{h}} = \frac{\text{SNR}_{ss}^{-1}}{8\pi^2 L^3\left(\sum_{k=0}^{U-1} i_k^2 - \frac{\left(\sum_{n=0}^{U-1} i_n\right)^2}{U}\right)} = \text{CRB}_{|\text{SNR}_{ss}}. \quad \text{(A-13)}$$

APPENDIX B

In this appendix, we will prove that there always exists a structure with an even number of training signal subblocks which has a smaller snapshot CRB than a structure with an odd number of training signal subblocks.

Consider three training signal structures with $(2K+2)$, $(2K+3)$, and $(2K+4)$ training signal subblocks. The corresponding best subblock location index vectors are given by

$$\boldsymbol{J}_{2K+2}^* = [0, 1, \ldots, K-1, P-K, P-K+1, \ldots, P-1] \quad \text{(B-1)}$$

$$\boldsymbol{J}_{2K+3}^* = [0, 1, \ldots, K, P-K, P-K+1, \ldots, P-1] \quad \text{(B-2)}$$

$$\boldsymbol{J}_{2K+4}^* = [0, 1, \ldots, K, P-K-1, P-K, \ldots, P-1]. \quad \text{(B-3)}$$

From (22), we obtain

$$\text{CRB}_{|\text{SNR}_i}^{-1}\left(\boldsymbol{J}_{2K+2}^*\right) = \mu\mathcal{E}_{2K+2}\left(\sum_{k=0}^{2K-1} i_k^2 - \frac{\left(\sum_{n=0}^{2K-1} i_n\right)^2}{2K}\right) \quad \text{(B-4)}$$

$$\mathrm{CRB}^{-1}_{|\mathrm{SNR}_i}\left(\boldsymbol{J}^*_{2K+3}\right) = \mu\mathcal{E}_{2K+3}\left(\sum_{k=0}^{2K} i_k^2 - \frac{\left(\sum_{n=0}^{2K} i_n\right)^2}{2K+1}\right) \tag{B-5}$$

$$\mathrm{CRB}^{-1}_{|\mathrm{SNR}_i}\left(\boldsymbol{J}^*_{2K+4}\right) = \mu\mathcal{E}_{2K+4}\left(\sum_{k=0}^{2K+1} i_k^2 - \frac{\left(\sum_{n=0}^{2K+1} i_n\right)^2}{2K+2}\right) \tag{B-6}$$

where $\mathcal{E}_{2K+2}$, $\mathcal{E}_{2K+3}$, and $\mathcal{E}_{2K+4}$ are the energy of a received training signal subblock for the structures $\boldsymbol{J}^*_{2K+2}$, $\boldsymbol{J}^*_{2K+3}$, and $\boldsymbol{J}^*_{2K+4}$, respectively, and $\mu = 8\pi^2L^2/\sigma_n^2$ is a constant. For a fixed transmitted training signal energy and a given channel, we have

$$\mathcal{E}_{2K+3} = \frac{2K+2}{2K+3}\mathcal{E}_{2K+2} \tag{B-7}$$

$$\mathcal{E}_{2K+4} = \frac{2K+2}{2K+4}\mathcal{E}_{2K+2}. \tag{B-8}$$

Substituting (B-7) and (B-8) into (B-4), (B-5), and (B-6) gives

$$\mathrm{CRB}^{-1}_{|\mathrm{SNR}_i}\left(\boldsymbol{J}^*_{2K+2}\right) = \mu\mathcal{E}_{2K+2}A_2 \tag{B-9}$$

$$\mathrm{CRB}^{-1}_{|\mathrm{SNR}_i}\left(\boldsymbol{J}^*_{2K+3}\right) = \mu\mathcal{E}_{2K+2}A_3 \tag{B-10}$$

$$\mathrm{CRB}^{-1}_{|\mathrm{SNR}_i}\left(\boldsymbol{J}^*_{2K+4}\right) = \mu\mathcal{E}_{2K+2}A_4 \tag{B-11}$$

where

$$A_2 = \frac{4K^3 - 6PK^2 + 3P^2K - K}{6} \tag{B-12}$$

$$A_3 = \frac{(2K+2)K}{3(2K+3)(2K+1)} \times (4K^3 - 6PK^2 + 3P^2K + 5K + 8K^2 - 9PK + 1 + 3P^2 - 3P) \tag{B-13}$$

$$A_4 = \frac{(K+1)^2\left(4(K+1)^2 - 6P(K+1) + 3P^2 - 1\right)}{6(K+2)}. \tag{B-14}$$

Let $P = \lambda K$ where $\lambda > 2$. Then we have

$$(A_2 - A_3) + (A_4 - A_3) = \frac{A\lambda^2 - B\lambda + C}{6(2K+3)(2K+1)(K+2)} = f(\lambda) \tag{B-15}$$

where

$$A = 6K^4 + 12K^3 + 9K^2$$
$$B = 12K^4 + 42K^3 + 54K^2 + 18K$$
$$C = 8K^4 + 40K^3 + 77K^2 + 44K + 9.$$

Since $A > 0$, $f(\lambda)$ is a convex function. Since

$$B^2 - 4AC = -[48K^8 + 336K^7 + 996K^6 + 1224K^5 + 672K^4 + 72K^3] < 0 \tag{B-16}$$

$f(\lambda)$ does not cross the zero line (the λ-axis of the $f(\lambda)$ plot) and we immediately obtain $f(\lambda) > 0$, i.e.,

$$(A_2 - A_3) + (A_4 - A_3) > 0. \tag{B-17}$$

Hence, we have either $(A_2 > A_3)$, or $(A_4 > A_3)$, or both A_2 and A_4 are larger than A_3. This result together with (B-9), (B-10), and (B-11) implies that there always exists a structure with an even number of training subblocks ($\boldsymbol{J}^*_{2K+2}$ or $\boldsymbol{J}^*_{2K+4}$) which has a smaller CRB than a structure with an odd number of training signal subblocks ($\boldsymbol{J}^*_{2K+3}$).

Appendix C

This appendix presents an analytical solution for $K^\star$. First we treat K as a continuous variable. After a straight-forward calculation of $df(K)/dK = 0$ yields

$$-8K^3 + 6(P-2)K^2 + 12PK - 3P^2 + 1 = 0. \tag{C-1}$$

Then the roots of the above cubic expression are obtained by Cardano's formula [24] as

$$K_1 = -\frac{A}{3} + (p+q) \tag{C-2}$$

$$K_2 = -\frac{A}{3} - \frac{1}{2}(p+q) + j\frac{\sqrt{3}}{2}(p-q) \tag{C-3}$$

$$K_3 = -\frac{A}{3} - \frac{1}{2}(p+q) - j\frac{\sqrt{3}}{2}(p-q) \tag{C-4}$$

where

$$A = \frac{6-3P}{4} \tag{C-5}$$

$$p = (d - b/2)^{1/3} \tag{C-6}$$

$$q = (-d - b/2)^{1/3} \tag{C-7}$$

$$b = \frac{-P^3 + 6P^2 + 12P + 4}{32} \tag{C-8}$$

$$d = j\frac{(P+1)\sqrt{6(P^3+2)}}{32}. \tag{C-9}$$

A direct calculation for the values of the three roots gives that $K_2 < 0$, $K_1 > P/2$, and $-1 < K_3 \le (3P-14)/20 < (P-1)/2$. Recall that the valid range of K is $[1, \lceil (P-1)/2 \rceil]$. Hence, we just need to consider K_3. Next, from the fact that $d^2f(K)/dK^2$ at $K = K_3$ is positive, we conclude that K_3 is a minimum point. If K_3 is within the valid range, $K^\star$ is either $\lfloor K_3 \rfloor$ or $\lceil K_3 \rceil$. If $-1 < K_3 < 1$, then $K^\star = 1$. Then $K^\star$ can be obtained as

$$K^\star = \underset{K \in \{\{1, \lfloor K_3 \rfloor, \lceil K_3 \rceil\} \cap \{1, \ldots, \lceil (P-1)/2 \rceil\}\}}{\arg\min} f(K) \tag{C-10}$$

which has been corroborated by the numerical result in Fig. 4.

Appendix D

This appendix derives the pdf of Z defined in (38). We can express Z as

$$Z = \boldsymbol{h}^H\boldsymbol{C}_{\boldsymbol{h}}^{-1/2}\boldsymbol{C}_{\boldsymbol{h}}^{1/2}\boldsymbol{S}_o^H\boldsymbol{S}_o\boldsymbol{C}_{\boldsymbol{h}}^{1/2}\boldsymbol{C}_{\boldsymbol{h}}^{-1/2}\boldsymbol{h} = \boldsymbol{h}^H\boldsymbol{C}_{\boldsymbol{h}}^{-1/2}\boldsymbol{U}\boldsymbol{\Omega}\boldsymbol{U}^H\boldsymbol{C}_{\boldsymbol{h}}^{-1/2}\boldsymbol{h} \tag{D-1}$$

$$= \boldsymbol{g}^H \boldsymbol{\Omega} \boldsymbol{g} = \sum_{k=0}^{L-1} |g_k|^2 \gamma_k \tag{D-2}$$

where $\boldsymbol{U}\boldsymbol{U}^H = \boldsymbol{I}_L$ and

$$\boldsymbol{g} = \boldsymbol{U}^H \boldsymbol{C}_{\boldsymbol{h}}^{-1/2} \boldsymbol{h} = [g_0, g_1, \ldots, g_{L-1}]^T \tag{D-3}$$

$$\boldsymbol{\Omega} = \mathrm{diag}\{\gamma_0, \gamma_1, \ldots, \gamma_{L-1}\}. \tag{D-4}$$

Note that independent zero-mean complex Gaussian random variables $\{h_k\}$ result in i.i.d. zero-mean complex Gaussian random variables $\{g_k\}$ with a variance of 1/2 per real dimension. The approach can be applied for correlated $\{h_k\}$, but for simplicity, we consider independent $\{h_k\}$. Suppose there are $m (1 \le m \le L)$ nonzero distinct values of γ_k denoted by $\{\lambda_l : l = 1, 2, \ldots, m\}$ and the number of γ_k's with the same value λ_l is κ_l. Define

$$\eta_l = \sum_{k:\gamma_k = \lambda_l} |g_k|^2. \tag{D-5}$$

Then $\{\eta_l\}$ are independent Chi-square random variables with $2\kappa_l$ DOFs. We have

$$Z = \sum_{l=1}^{m} \lambda_l \eta_l \tag{D-6}$$

and from [25] we obtain the pdf of Z as

$$p_Z(z) = \sum_{l=1}^{m} \sum_{k=1}^{\kappa_l} A_{lk} p_{lk}(z) \tag{D-7}$$

where

$$p_{lk}(z) = \frac{1}{\lambda_l} f_{2k}(z/\lambda_l) \tag{D-8}$$

$$f_n(x) = \frac{x^{\frac{n}{2}-1} e^{-x/b}}{b^{n/2}\Gamma(n/2)},\ b = \sigma_{g_k}^2 = 1;\ 0 < x < \infty \tag{D-9}$$

$$A_{lk} = \left\{ \prod_{i=1}^{m} (-2\lambda_i)^{-\kappa_i} \right\} (-1)^k (2\lambda_l)^k b_{lk} \tag{D-10}$$

$$b_{lk} = \left\{ \sum_{j_1=0}^{\kappa_l - k - 1} \binom{\kappa_l - k - 1}{j_1} A_l^{(\kappa_l - k - 1 - j_1)} \times \sum_{j_2=0}^{j_1 - 1} \binom{j_1 - 1}{j_2} A_l^{(j_1 - 1 - j_2)} \cdots \right\} \frac{\Delta_l}{(\kappa_l - k)!} \tag{D-11}$$

$$\Delta_l = \prod_{k=1, k\neq l}^{m} \left(\frac{1}{2\lambda_l} - \frac{1}{2\lambda_k} \right)^{-\kappa_l} \tag{D-12}$$

$$A_l^{(n)} = (-1)^{(n+1)} n! \times \sum_{k=1, k\neq l}^{m} \kappa_k \left(\frac{1}{2\lambda_l} - \frac{1}{2\lambda_k} \right)^{-(n+1)}. \tag{D-13}$$

APPENDIX E

In this appendix, we derive σ_Z^2, the variance of the energy of a received training subblock. Equation (38) can be expressed as

$$Z = \mathrm{trace}\left[\boldsymbol{S}_o \boldsymbol{h}\boldsymbol{h}^H \boldsymbol{S}_o^H \right] \tag{E-1}$$

and the first and second moments of Z are given by

$$E[Z] = \mathrm{trace}\left[\boldsymbol{S}_o \boldsymbol{C}_{\boldsymbol{h}} \boldsymbol{S}_o^H \right] = \sum_{k=0}^{L-1} |s_k|^2 = \mathcal{E}_{\mathrm{T1}} \tag{E-2}$$

$$\begin{aligned} E[Z^2] &= E\left[\boldsymbol{h}^H \boldsymbol{S}_o^H \boldsymbol{S}_o \boldsymbol{h}\boldsymbol{h}^H \boldsymbol{S}_o^H \boldsymbol{S}_o \boldsymbol{h} \right] \\ &= E\left[\mathrm{trace}[\boldsymbol{C}_{\boldsymbol{S}} \boldsymbol{h}\boldsymbol{h}^H \boldsymbol{C}_{\boldsymbol{S}} \boldsymbol{h}\boldsymbol{h}^H] \right] \\ &= E\left[\mathrm{trace}\left[\boldsymbol{C}_{\boldsymbol{S}} \widetilde{\boldsymbol{C}}_h \boldsymbol{C}_{\boldsymbol{S}} \widetilde{\boldsymbol{C}}_h \right] \right] \end{aligned} \tag{E-3}$$

where $\boldsymbol{C}_{\boldsymbol{S}} = \boldsymbol{S}_o^H \boldsymbol{S}_o$ and $\widetilde{\boldsymbol{C}}_{\boldsymbol{h}} = \boldsymbol{h}\boldsymbol{h}^H$. After some manipulations, (E-3) can be expressed as

$$E[Z^2] = \sum_{l=0}^{L-1} \sum_{k=0}^{L-1} \sum_{m=0}^{L-1} \sum_{n=0}^{L-1} \boldsymbol{C}_{\boldsymbol{S}}(k,m) \boldsymbol{C}_{\boldsymbol{S}}(l,n) E \times \left[\widetilde{\boldsymbol{C}}_h(m,l) \widetilde{\boldsymbol{C}}_h(n,k) \right] \tag{E-4}$$

where

$$E\left[\widetilde{\boldsymbol{C}}_h(m,l) \widetilde{\boldsymbol{C}}_h(n,k) \right] = \begin{cases} 2\sigma_{h_k}^4, & \text{if } m=l=n=k \\ \sigma_{h_m}^2 \sigma_{h_n}^2, & \text{if } m=l, n=k, m\neq n \\ \sigma_{h_m}^2 \sigma_{h_l}^2, & \text{if } m=k, l=n, m\neq l \\ 0 & \text{else.} \end{cases} \tag{E-5}$$

Then, from (E-2), (E-4) and (E-5), we obtain the variance of Z as

$$\begin{aligned} \sigma_Z^2 = &\sum_{k=0}^{L-1} \left(\boldsymbol{C}_{\boldsymbol{S}}(k,k) \right)^2 2\sigma_{h_k}^4 \\ &+ \sum_{m=0}^{L-1} \sum_{n=0, n\neq m}^{L-1} \boldsymbol{C}_{\boldsymbol{S}}(m,n) \boldsymbol{C}_{\boldsymbol{S}}(n,m) \sigma_{h_m}^2 \sigma_{h_n}^2 \\ &+ \sum_{i=0}^{L-1} \sum_{l=0, l\neq i}^{L-1} \boldsymbol{C}_{\boldsymbol{S}}(i,i) \boldsymbol{C}_{\boldsymbol{S}}(l,l) \sigma_{h_i}^2 \sigma_{h_l}^2 - \mathcal{E}_{\mathrm{T1}}^2. \end{aligned} \tag{E-6}$$

APPENDIX F

This appendix presents the MSE of the LS channel estimation ($\mathrm{MSE}_{\boldsymbol{h}}$) using an arbitrary training signal. We assume a perfect frequency synchronization (or a negligible residual frequency offset after frequency offset estimation and compensation). After frequency offset compensation, the observation vector of length UL for the LS chanel estimation is given by

$$\bar{\boldsymbol{y}} = \boldsymbol{\Gamma}^H(v) \boldsymbol{y} = \boldsymbol{C}\boldsymbol{h} + \bar{\boldsymbol{w}} \tag{F-1}$$

where $\bar{\boldsymbol{w}}$ is a zero-mean, circularly symmetric, complex Gaussian noise vector with a covariance matrix of $\sigma_n^2 \bar{\boldsymbol{I}}_{UL}$. The LS channel estimate is given by

$$\hat{\boldsymbol{h}} = (\boldsymbol{C}^H \boldsymbol{C})^{-1} \boldsymbol{C}^H \bar{\boldsymbol{y}} \tag{F-2}$$

$$= \left(\boldsymbol{S}_o^H \boldsymbol{S}_o\right)^{-1} \boldsymbol{S}_o^H \boldsymbol{S}_o \boldsymbol{h} + \frac{1}{U}\left[\boldsymbol{1}_U^T \otimes \left(\left(\boldsymbol{S}_o^H \boldsymbol{S}_o\right)^{-1} \boldsymbol{S}_o^H\right)\right] \bar{\boldsymbol{w}} \tag{F-3}$$

$$= \boldsymbol{F}_L^H \boldsymbol{\Psi} \boldsymbol{F}_L \boldsymbol{h} + \frac{1}{U}\left[\boldsymbol{1}_U^T \otimes \left(\boldsymbol{F}_L^H \boldsymbol{\Psi} \boldsymbol{F}_L\right)\right] \bar{\boldsymbol{w}} \tag{F-4}$$

where a pseudo-inverse is applied if the training matrix $\boldsymbol{S}_o$ is rank-deficient. The channel estimation MSE is obtained by

$$\mathrm{MSE}_{\boldsymbol{h}} = E\left[\|\hat{\boldsymbol{h}} - \boldsymbol{h}\|^2\right] \tag{F-5}$$

$$= \mathrm{trace}\left[\boldsymbol{F}_L^H (\boldsymbol{I}_{UL} - \boldsymbol{\Psi}) \boldsymbol{F}_L \boldsymbol{C}_{\boldsymbol{h}} \boldsymbol{F}_L^H (\boldsymbol{I}_{UL} - \boldsymbol{\Psi}) \boldsymbol{F}_L\right] + \frac{\sigma_n^2}{U} \sum_{i:\Sigma_i \neq 0} \frac{1}{|\Sigma_i|^2} \tag{F-6}$$

$$= \frac{N_{\mathrm{null}}}{L} + \frac{\sigma_n^2}{U} \sum_{i:\Sigma_i \neq 0} \frac{1}{|\Sigma_i|^2} \tag{F-7}$$

where Σ_i represents i-th diagonal element of $\boldsymbol{\Sigma}$ from (52) and N_{null} is the number of Σ_i's having zero value. For a full-rank $\boldsymbol{S}_o$, the first term in (F-7) vanishes. Note that (F-7) is minimized when all Σ_i's are constant which corresponds to the MSE-minimizing optimal training signal for the LS channel estimation.

APPENDIX G

This appendix presents how to obtain a robust K value for the proposed training structure. Applying the method of [18] on our training structure gives the following maximum-likelihood frequency offset estimator:

$$\hat{v} = \underset{\tilde{v} \in R_v}{\arg\max}\, \mathcal{M}(\tilde{v}) \tag{G-1}$$

where R_v is the search range of v and $\mathcal{M}(\tilde{v})$ is the likelihod metric defined by

$$\mathcal{M}(\tilde{v}) = \boldsymbol{y}^H \boldsymbol{\Gamma}(\tilde{v}) \boldsymbol{D} \boldsymbol{\Gamma}^H(\tilde{v}) \boldsymbol{y}. \tag{G-2}$$

The proposed training structure with optimal K value gives a better snapshot CRB and average CRB than other K values, but gives a better MSE only at moderate and high SNR. At low SNR, the optimal K value yields more outliers due to large sidelobes of the likelihood metric. We can design a robust K value for which the sidelobes of its likelihood metric are less than some threshold. We find that the threshold of about 0.75 of the mainlobe peak would be a good design choice. For simplicity in the robust K design, we can replace $\boldsymbol{y}$ with the corresponding transmitted training signal $(\boldsymbol{1}_{2K} \otimes \boldsymbol{s}_o)$ where $\boldsymbol{s}_o = [s_0, s_1, \ldots, s_{L-1}]^T$ is the transmitted subblock signal. By plotting the normalized likelihood metric $\mathcal{M}(\tilde{v})/\max\{\mathcal{M}(\tilde{v})\}$ for different K values, we can obtain a robust K value. As an example, the normalized likelihood metrics for the proposed structure with $P = 39$, $K = 4$ and $K = 9$, and the conventional structure with $K_c = 9$ and $K_c = 19$ are shown in Fig. 18. The proposed structure has a sharper mainlobe than the conventional structure, but also has larger sidelobes. We notice that $K = 4$ does not satisfy the robust K design criterion, but $K = 9$ does.

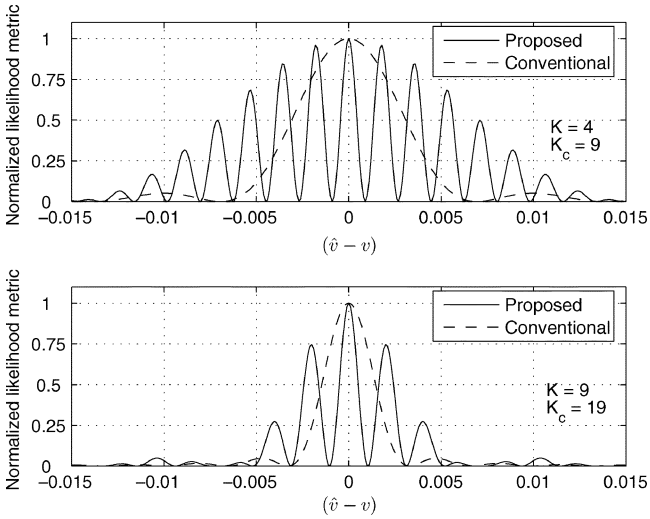


Fig. 18. Normalized likelihood metric of frequency offset estimation for the proposed training structure and the conventional structure using the same system resource (i.e., $K_c = 2K + 1$).

ACKNOWLEDGMENT

H. Minn acknowledges Mr. S. Xing for performing earlier simulation and part of the calculation for Appendix A. The authors thank the reviewers for their valuable comments.

REFERENCES

[1] A. Milewski, "Periodic sequences with optimal properties for channel estimation and fast start-up equalization," *IBM J. Res. Dev.*, vol. 27, no. 5, pp. 426–431, Sep. 1983.

[2] S. N. Crozier, D. D. Falconer, and S. A. Mahmoud, "Least sum of squared errors (LSSE) channel estimation," *Inst. Elect. Eng. Proc.-F*, vol. 138, no. 4, pp. 371–378, Aug. 1991.

[3] C. Tellambura, M. G. Parker, Y. J. Guo, S. J. Shepherd, and S. K. Barton, "Optimal sequences for channel estimation using discrete Fourier transform techniques," *IEEE Trans. Commun.*, vol. 47, no. 2, pp. 230–238, Feb. 1999.

[4] W. Chen and U. Mitra, "Training sequence optimization: Comparison and an alternative criterion," *IEEE Trans. Commun.*, vol. 48, no. 12, pp. 1987–1991, Dec. 2002.

[5] D. C. Chu, "Polyphase codes with good periodic correlation properties," *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 531–532, Jul. 1972.

[6] W. H. Mow, "A unified construction of perfect polyphase sequences," in *Proc. IEEE Int. Symp. Inf. Theory*, 1995, p. 459.

[7] ——, "A new unified construction of perfect root-of-unity sequences," in *Proc. Spread Spectrum Tech. Appl.*, Mainz, Germany, 1996, pp. 955–959.

[8] J. C. L. Ng, K. B. Letaief, and R. D. Murch, "Complex optimal sequences with constant magnitude for fast channel estimation initialization," *IEEE Trans. Commun.*, vol. 46, no. 3, pp. 305–308, Mar. 1998.

[9] M. Dong and L. Tong, "Optimal design and placement of pilot symbols for channel estimation," *IEEE Trans. Signal Process.*, vol. 50, no. 12, pp. 3055–3069, Dec. 2002.
[10] X. Ma, G. B. Giannakis, and S. Ohno, "Optimal training for block transmissions over doubly selective wireless fading channels," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1351–1365, May 2003.
[11] H. Minn and N. Al-Dhahir, "Optimal training signals for MIMO OFDM channel estimation," in *Proc. IEEE Globecom*, Nov. 29–Dec. 3, 2004, vol. 1, pp. 219–224.
[12] R. Dabora, J. Goldberg, and H. Messer, "Inherent limitations in data-aided timing synchronization of continuous phase modulation signals over time-selective fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1470–1482, Jun. 2002.
[13] H. Minn, V. K. Bhargava, and K. B. Letaief, "A robust timing and frequency synchronization for OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 822–839, Jul. 2003.
[14] H. Minn and S. Xing, "An optimal training signal structure for frequency offset estimation," *IEEE Trans. Commun.*, vol. 53, no. 2, pp. 343–355, Feb. 2005.
[15] P. Ciblat and L. Vandendorpe, "On the maximum-likelihood based data-aided frequency offset and channel estimates," in *Proc. EURASIP Eur. Signal Process. Conf.*, Toulouse, France, Sep. 2002, vol. 1, pp. 627–630.
[16] O. Besson and P. Stoica, "Data-aided frequency offset estimation in frequency selective channels: Training sequence selection," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2002, vol. 3, pp. 2445–2448.
[17] P. Stoica and O. Besson, "Training sequence design for frequency offset and frequency-selective channel estimation," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1910–1917, Nov. 2003.
[18] M. Morelli and U. Mengali, "Carrier-frequency estimation for transmissions over selective channels," *IEEE Trans. Commun.*, vol. 48, no. 9, pp. 1580–1589, Sep. 2000.
[19] F. Gini and R. Reggiannini, "On the use of Cramer–Rao-like bounds in the presence of random nuisance parameters," *IEEE Trans. Commun.*, vol. 48, no. 12, pp. 2120–2126, Dec. 2000.
[20] Q. T. Zhang and D. P. Liu, "A simple capacity formula for correlated diversity Rician fading channels," *IEEE Commun. Lett.*, vol. 6, no. 11, pp. 481–483, Nov. 2002.
[21] J. Li, G. Liu, and G. B. Giannakis, "Carrier frequency offset estimation for OFDM-based WLANs," *IEEE Signal Process. Lett.*, vol. 8, no. 3, pp. 80–82, Mar. 2001.
[22] H. Minn, P. Tarasak, and V. K. Bhargava, "OFDM frequency offset estimation based on BLUE principle," in *Proc. IEEE Veh. Technol. Conf.*, Vancouver, BC, Canada, Sep. 2002, pp. 1230–1234.
[23] H. Minn and V. K. Bhargava, "A reduced complexity, improved frequency offset estimation for OFDM-based WLANs," in *Proc. WNCG Wireless Networking Symp.*, Austin, TX, Oct. 2003.
[24] A. Jeffrey, *Handbook of Mathematical Formulas and Integrals*, 2nd ed. New York: Academic, 2000.
[25] A. M. Mathai and S. B. Provost, *Quadratic Forms in Random Variables: Theory and Applications*. New York: Marcel Dekker, 1992.
[26] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: High-Speed Physical Layer in the 5 GHz Band*, IEEE Standard 802.11a, IEEE LAN/MAN Standards Committee, 1999.

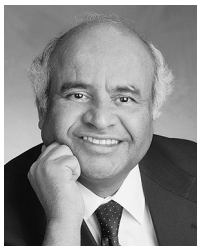
Hlaing Minn (S'99–M'01) received the B.E. degree in electronics from Yangon Institute of Technology, Yangon, Myanmar, in 1995, the M.Eng. degree in telecommunications from Asian Institute of Technology (AIT), Pathumthani, Thailand, in 1997, and the Ph.D. degree in electrical engineering from the University of Victoria, Victoria, BC, Canada, in 2001.

He was with the Telecommunications Program in AIT as a Laboratory Supervisor during 1998. He was a Research Assistant from 1999 to 2001 and a Postdoctoral Research Fellow during 2002 in the Department of Electrical and Computer Engineering at the University of Victoria. Since September 2002, he has been with the Erik Jonsson School of Engineering and Computer Science, the University of Texas at Dallas, Richardson, as an Assistant Professor. His research interests include wireless communications, statistical signal processing, error control, detection, estimation, synchronization, signal design, and cross-layer design.

Dr. Minn is an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS.

Xiaoyu Fu (S'04) received the B.S. and M.S. degrees in electrical engineering from the University of Electronic Science and Technology of China, Sichuan, China, in 1999 and 2002, respectively. He is currently working toward the Ph.D. degree at the University of Texas at Dallas (UTD), Richardson.

Since 2003, he has been a Research Assistant with the Information and Communication Systems Laboratory at the Department of Electrical Engineering, UTD. His research interests include synchronization and channel estimation in wireless communications systems.



Vijay K. Bhargava (S'70–M'74–SM'82–F'92) received the B.Sc., M.Sc., and Ph.D. degrees from Queen's University, Kingston, ON, Canada in 1970, 1972, and 1974, respectively.

Currently, he is a Professor and Head of the Department of Electrical and Computer Engineering at the University of British Columbia, Vancouver, BC, Canada. Previously, he was with the University of Victoria (1984–2003) and with Concordia University in Montréal (1976–1984). He is a coauthor of the book *Digital Communications by Satellite* (New York: Wiley, 1981) and co-editor of the books *Reed–Solomon Codes and Their Applications* (Piscataway, NJ: IEEE Press, 1994) and *Communications, Information and Network Security* (Boston: Kluwer, 2003). His research interests are in multimedia wireless communications.

Dr. Bhargava is a Fellow of the British Columbia Advanced Systems Institute, the Engineering Institute of Canada (EIC), the Canadian Academy of Engineering, and the Royal Society of Canada. He is a recipient of the IEEE Centennial Medal (1984), IEEE Canada's McNaughton Gold Medal (1995), the IEEE Haraden Pratt Award (1999), the IEEE Third Millennium Medal (2000), the IEEE Graduate Teaching Award (2002), and the Eadie Medal of the Royal Society of Canada (2004). He has served on the Boards of the IEEE Information Theory and Communications Societies, and is a Past President of the IEEE Information Theory Society. He is an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.