An Optimal Training Signal Structure for Frequency-Offset Estimation

Hlaing Minn, *Member, IEEE,* and Shaohui Xing

Abstract—This paper addresses an optimal training-signal design for frequency-offset estimation. Based on minimizing the Cramer–Rao lower bound for frequency-offset estimation with constraints on the peak and the total training signal energies, and the training block length, the optimal training-signal structure is developed. An approximate version of the optimal training-signal structure is proposed, which has practically the same performance as the optimal one, and provides convenience in training-signal generation and estimator derivation. Two robust reduced-complexity frequency-offset estimation methods for the proposed training structures are presented. In order to handle larger frequency offsets, modified training-signal structures are proposed. Frequency-offset estimation methods suitable for these training signals are also derived, based on the best linear unbiased estimation principle. Analytical and simulation results show that the proposed training-signal structures improve the estimation performance significantly.

Index Terms—Best linear unbiased estimation (BLUE), Cramer–Rao bound (CRB), frequency-offset estimation, peak-to-average sample energy ratio (PAR), training design.

I. INTRODUCTION

TRAINING signals are often used in communications systems for timing synchronization, frequency synchronization, and channel estimation. In [1] and [2] (and references therein), training-signal design for channel estimation in single-carrier systems was discussed. In [3], training-signal design for channel estimation in orthogonal frequency-division multiplexing (OFDM) systems was described. In [4] and [5], some training-signal designs for timing synchronization in OFDM systems were prsesented. For frequency synchronization in OFDM systems, [6] used a training signal consisting of two identical parts $(L = 2)$, while [7]–[9] used a training signal composed of $L > 2$ identical parts. These training signals are generated by the inverse fast Fourier transform (IFFT) of a pseudonoise (PN) sequence on every Lth subcarrier in the frequency domain. The above timing- and frequency-synchronization methods are based on time-domain processing, and hence, also applicable to single-carrier systems with the same training signal. Alternatively, the training signal with L

identical parts can be constructed such that each part is of a time-domain PN sequence.

In this paper, we investigate training signal design for frequency-offset estimation in an additive white Gaussian noise (AWGN) channel. Based on minimizing the Cramer–Rao lower bound (CRB) for frequency-offset estimation with constraints on the peak and the total training-signal energies, and the training block length, the optimal training-signal structure is developed. The proof for the optimality and uniqueness of the proposed training-signal structure is presented. Then, an approximate version of the optimal training-signal structure is proposed, which has practically the same performance as the optimal one, and provides convenience in training-signal generation and estimator derivation. The maximum-likelihood (ML) estimator from [13] can be appropriately applied to the proposed training structures with a large number of nonzero training samples, but it fails to provide a reliable estimate when the number of the nonzero training samples in the proposed training structures is small and the signal-to-noise ratio (SNR) is not high. Furthermore, the complexity of the ML estimator is quite high. To reduce the estimation complexity, two frequency-offset estimation methods are presented, which are also robust to the parameters of the proposed training structures. To handle larger frequency offsets, modified training signal structures are proposed, and the corresponding estimation methods are derived based on the best linear unbiased estimation (BLUE) principle. Analytical and simulation results show that the proposed training-signal structures improve the estimation performance significantly.

The rest of the paper is organized as follows. Section II derives CRBs, and Section III presents the optimal training-signal structure, while the proof of the optimality and the uniqueness of the proposed training-signal structure is provided in the Appendix. Frequency-offset estimators for the proposed training-signal structures are addressed in Section IV. Analytical and simulation results and estimators' complexities are discussed in Section V, and the paper is concluded in Section VI.

II. CRAMER–RAO LOWER BOUND (CRB)

Frequency offset is unavoidable at the receiver due to the oscillators' inaccuracies and the Doppler shifts of the mobile wireless channel. To eliminate or reduce the performance degradation caused by the frequency offset, the receiver usually performs frequency synchronization or frequency-offset compensation based on the frequency-offset estimation. In this paper, we consider frequency-offset estimation based on the training

The authors are with the Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX 75083-0688 USA (e-mail: hlaing.minn@utdallas.edu; shaohui@utdallas.edu).

signal vector $s = [s_0, s_1, \ldots, s_{N-1}]^T$. Let us consider a signal model given by

$$r = W(v)s + n \tag{1}$$

where r is the received training vector of length N, $W(v)$ is a diagonal matrix with diagonal elements $\{\exp(j2\pi 0v/N), \exp(j2\pi 1v/N), \exp(j2\pi 2v/N), \ldots, \exp(j2\pi(N-1)v/N)\}$ corresponding to the normalized frequency offset v (normalized by $1/(NT)$ where T is the sample duration), and n is the zero-mean, complex Gaussian noise vector with a covariance matrix $C_n = \sigma_n^2 I$, where I is an identity matrix.

For the parameter v, the received vector r has a complex Gaussian probability density function (pdf) $p(r; v)$, with the mean vector and the covariance matrix, respectively, given by $\mu_r = W(v)s$ and $C_r = C_n = \sigma_n^2 I$. The Fisher information in this case is given by

$$J = E\left[\frac{\partial \ln p(r; v)}{\partial v}\frac{\partial \ln p(r; v)}{\partial v}\right] \tag{2}$$

$$= tr\left[C_r^{-1}\frac{\partial C_r}{\partial v}C_r^{-1}\frac{\partial C_r}{\partial v}\right]$$

$$+ 2\Re\left[\frac{\partial \mu_r^H}{\partial v}C_r^{-1}\frac{\partial \mu_r}{\partial v}\right] \tag{3}$$

where $E[\cdot]$ represents the expectation, $tr[\cdot]$ is the trace, $\Re[\cdot]$ denotes the real part, and the superscript H is the Hermitian transpose. After some calculation, we have

$$J = \frac{8\pi^2 s^H \Lambda^2 s}{N^2 \sigma_n^2} \tag{4}$$

where $\Lambda = \text{diag}\{0, 1, \ldots, N-1\}$ is a diagonal matrix with diagonal elements $\{0, 1, \ldots, N-1\}$. The CRB of the frequency-offset estimation for the signal model (1) is given by

$$\text{CRB}_{np} = J^{-1} = \frac{N^2 \sigma_n^2}{8\pi^2 s^H \Lambda^2 s} \tag{5}$$

where the subscript np indicates that the corresponding signal model includes no phase noise.

Next, let us consider a signal model containing an arbitrary carrier phase ψ as

$$r = e^{j\psi}W(v)s + n. \tag{6}$$

For the parameter vector $\alpha = [v, \psi]^T$, the received vector r has a complex Gaussian pdf, $p(r; \alpha)$, with the mean vector and the covariance matrix, respectively, given by $\mu_r = e^{j\psi}W(v)s$ and $C_r = C_n = \sigma_n^2 I$. The $[i, j]$ element of the Fisher information matrix J is given by

$$J_{i,j} = E\left[\frac{\partial \ln p(r; \alpha)}{\partial \alpha_i}\frac{\partial \ln p(r; \alpha)}{\partial \alpha_j}\right] \tag{7}$$

$$= tr\left[C_r^{-1}\frac{\partial C_r}{\partial \alpha_i}C_r^{-1}\frac{\partial C_r}{\partial \alpha_j}\right]$$

$$+ 2\Re\left[\frac{\partial \mu_r^H}{\partial \alpha_i}C_r^{-1}\frac{\partial \mu_r}{\partial \alpha_j}\right] \tag{8}$$

where α_i is the ith element of α. After some calculation, we have

$$J_{1,1} = \frac{8\pi^2}{N^2 \sigma_n^2}s^H \Lambda^2 s \tag{9}$$

$$J_{1,2} = J_{2,1} = \frac{4\pi}{N\sigma_n^2}s^H \Lambda s \tag{10}$$

$$J_{2,2} = \frac{2}{\sigma_n^2}s^H s. \tag{11}$$

The CRB of the frequency-offset estimation for the signal model (6) is given by

$$\text{CRB} = [J^{-1}]_{1,1} = \left(J_{1,1} - J_{1,2}J_{2,2}^{-1}J_{2,1}\right)^{-1} \tag{12}$$

$$= \frac{N^2 \sigma_n^2}{8\pi^2}\frac{s^H s}{s^H s s^H \Lambda^2 s - (s^H \Lambda s)^2}. \tag{13}$$

III. TRAINING-SIGNAL DESIGN

In this section, we investigate the training-signal design that minimizes the CRB. The best training vector $s^\dagger$ is given by

$$s^\dagger = \arg\min_s \{\text{CRB}\}, \quad \text{constraint: } s^H s = E_s \tag{14}$$

where E_s is the total energy of the training-signal vector.

First, let us consider CRB_{np} (5) which corresponds to the signal model (1). We have

$$s^\dagger = \arg\max_s \left(s^H \Lambda^2 s\right), \quad \text{constraint}: s^H s = E_s \tag{15}$$

$$= \arg\max_s \left(\sum_{k=0}^{N-1}|s_k|^2 k^2\right), \quad \text{constraint}: s^H s = E_s. \tag{16}$$

From the above equation, it is straightforward to see that the best training-signal vector is determined by

$$\text{abs}(s^\dagger) = [0, 0, \ldots, 0, \sqrt{E_s}]^T \tag{17}$$

where $\text{abs}(\cdot)$ means the absolute values, and the superscript T denotes the transpose. Similarly, the worst training-signal vector $s^\sharp$ is given by

$$\text{abs}(s^\sharp) = [\sqrt{E_s}, 0, \ldots, 0]^T. \tag{18}$$

The solution (17) suggests that we need to concentrate all the training energy only at the last training sample. Some issues are involved in this solution. First of all, it suggests transmitting only one training sample preceded by null samples. In practice, it may be treated as the first training sample, and hence, it becomes the worst design (18). Secondly, even if v is small enough that no ambiguity will result, there is always an arbitrary carrier phase (phase noise) which will completely destroy the frequency-offset estimation based on one training sample. These facts indicate that in designing the training signal, the signal model should include an arbitrary carrier phase.

Now, let us consider CRB (13) corresponding to the signal model (6). From (14), and after some manipulation, we have the best training signal given by

$$s^{\dagger} = \arg\min_{s} \left((s^H \Lambda s)^2 - s^H s s^H \Lambda^2 s \right),$$
$$\text{constraint} : s^H s = E_s \tag{19}$$
$$= \arg\min_{s} \left(\sum_{k=0}^{N-1} \sum_{n=0}^{N-1} |s_k|^2 |s_n|^2 (nk - n^2) \right),$$
$$\text{constraint} : s^H s = E_s. \tag{20}$$

From (20), it can be observed that the phases of the training samples have no effect on the CRB. It is only the training-energy allocation among the training samples that has an effect on the CRB, and that has to be found. Define the following:

$$y = [|s_0|^2, |s_1|^2, \ldots, |s_{N-1}|^2]^T \tag{21}$$
$$G_{k,n} = n(k - n), 0 \le k \le (N - 1),$$
$$0 \le n \le (N - 1) \tag{22}$$

where $G_{k,n}$ is the $[k, n]$ element of an $N \times N$ matrix G. Note that the matrix G is singular and not symmetric. y can be viewed as the training-energy vector or the training-energy allocation. Then the best training vector is determined by the best training-energy allocation given by

$$y^{\dagger} = \arg\min_{y} \left\{ y^T G y \right\},$$
$$\text{constraint} : y^T \mathbf{1} = E_s \text{ and } 0 \le y_i \le P \tag{23}$$

where $\mathbf{1}$ is an all-one vector of length N, and P is a design value representing the allowable peak training-sample energy. The reason for the introduction of the upper limit P is to avoid nonlinear distortion of the training signal at the transmit amplifier [and also at the digital-to-analog converter (DAC)]. The optimization (23) containing equality and nonequality constraints can be solved by quadratic programming (QP) such as MATLAB function *quadprog.m*. The general result can be given by (24), shown at the bottom of the page.

However, since the matrix G is singular and not symmetric, the optimality and uniqueness of the above solution is uncertain at this point. We presented another approach for obtaining a solution to (23) in [10], which also gives the same result as QP does. Due to the optimization for the dominant terms only, at some steps, the approach from [10] does not guarantee the optimality and uniqueness of the solution in (24). Hence, we investigate the optimality and uniqueness of the solution in (24), and find that it is optimal and unique. The proof is presented in Appendix A for conciseness.

IV. FREQUENCY-OFFSET ESTIMATION

In this section, we consider frequency-offset estimation methods suitable for the proposed optimal training-signal structure. Consider the constraints of the total training-signal energy E_s, the training block length of N samples, and the allowable peak training-sample energy P. Note that $E_s \le PN$, otherwise the constraints cannot be satisfied. Let $K = \lceil E_s/(2P) \rceil$. Then the optimal training signal has a total of $2K$ nonzero samples with the sample indexes $[0, 1, \ldots, K-1, N-K, N-K+1, \ldots, N-1]$. All nonzero samples have equal sample energy of P, except the two samples at indexes $K-1$ and $N-K$, which have an equal sample energy given by $(E_s - 2(K-1)P)/2$, $(\le P)$. Data signals can be transmitted between the two nonzero training parts.

First, we consider the ML estimator (MLE#1) from [13]. In estimation, the data signals between the two nonzero training parts are replaced with zeros. After that, MLE#1 can be directly applied. We find that MLE#1 gives mean-square error (MSE) almost equal to CRB for the proposed optimal training structure with a large number of nonzero training samples. However, MLE#1 fails to provide a reliable estimate when the number of nonzero training samples in the optimal training structure is small and the SNR is not high. Details will be discussed in the simulation section. Furthermore, the complexity of MLE#1 is quite high. Hence, in the following, we pursue reduced-complexity frequency-offset estimation methods which are also robust to the parameters of the optimal training structure.

Consider an approximate version of the optimal training signal where all $2K$ nonzero training samples have equal energy of $\tilde{P} = E_s/(2K) \le P$. If $K = E_s/(2P)$, then $\tilde{P} = P$, and the approximate version is identical to the optimal one. For practical systems, where K is not small, there is virtually no performance loss if the optimal training signal is replaced with its approximate version. This fact will be seen in the simulation results and discussions section. Hence, for simplicity in practice, one can use the approximate version of the optimal training signal.

In the following, we derive a few frequency-offset estimation methods based on the approximate version of the optimal training signal. The estimation methods for the optimal training signal can be similarly derived. But it requires a more lengthy calculation for not so small K, and would not be of practical interest, since the performance of the optimal training signal is practically not different from that of the approximate version.

First, we derive a frequency-offset estimation method based on sample-wise correlation and the BLUE principle. It will be denoted as S-BLUE. Next, we derive another method based on part-wise correlation and BLUE. It will be denoted as P-BLUE.

$$y_k^{\dagger} = \begin{cases} \min\left\{P, \frac{E_s}{2}\right\}, & \text{if } k = 0 \text{ or } N - 1 \\ \min\left\{P, \frac{E_s}{2} - \sum_{l=0}^{m-1} y_l^{\dagger}\right\}, & \text{if } k = m \text{ or } k = N - 1 - m; 1 \le m \le \frac{N}{2} - 1 \\ E_s - 2\sum_{l=0}^{k-1} y_l^{\dagger}, & \text{if } k = \frac{(N-1)}{2} \text{ and } N \text{ is an odd integer.} \end{cases} \tag{24}$$

The estimation range of S-BLUE is $|v| < N/(2(N-2))$, and that of P-BLUE-2 (where 2 indicates two nonzero parts) is $|v| < N/(2N - 2K)$. To extend the estimation range of P-BLUE, we propose a modified training signal which contains three or four nonzero parts. The frequency-offset estimation methods for the modified training signals with three and four nonzero parts are then derived, which will be denoted as P-BLUE-3 and P-BLUE-4, respectively.

For the same total training-signal energy E_s and the same peak-sample energy $\tilde{P} \leq P$, the number of samples in a nonzero part is given by $K = E_s/(n\tilde{P})$ for the training signal with $n(= 2, 3)$ nonzero parts. Assuming a nondispersive channel, we can multiply the received training samples with the conjugate of the transmitted training samples to get the same phase. Hence, for simplicity, all nonzero training samples are assumed to have the same phase.

A. S-BLUE Method

Performing sample-wise correlation, we obtain several correlation terms $\{R(m)\}$ given in (25) or (26), as shown at the bottom of the page, with correlation distance $m \in D$. If $N \geq 3K - 1$, we have $D = [1, 2, \ldots, K-1, N-2K+1, N-2K+2, \ldots, N-1]$, and $R(m)$ is given by (25). If $N \geq 3K - 2$, $D = [1, 2, \ldots, N-1]$ and $R(m)$ is given by (26). For $N \geq 3K - 1$, we have $D = [1, 2, \ldots, K-1, N-2K+1, N-2K+2, \ldots, N-1]$, and for $N \leq 3K - 2$, $D = [1, 2, \ldots, N-1]$. Then we have several estimates $\{\theta_k\}$ given by

$$\theta_k = \frac{N}{2\pi D_k} \angle\{R(D_k)\} \qquad (27)$$

where D_k is the kth element of D, and $\angle\{X\}$ represents the angle of X. Let $\boldsymbol{\theta}$ denote the column vector of $\{\theta_k\}$. Then the frequency-offset estimate based on BLUE can be given by

$$\hat{v} = \boldsymbol{\theta}^T \boldsymbol{w} \qquad (28)$$

where the weighting vector $\boldsymbol{w}$ is calculated as follows [11]:

$$\boldsymbol{w} = \frac{C_\theta^{-1} \mathbf{1}}{\mathbf{1}^T C_\theta^{-1} \mathbf{1}}. \qquad (29)$$

The calculation of C_θ for S-BLUE is lengthy, and the results are given in Appendix B. Since the maximum correlation distance is $N - 1$, the estimation range of S-BLUE is limited by $|v| < N/(2N - 2)$.

B. P-BLUE-2 Method

In the BLUE method with part-wise correlation between two nonzero parts, the phases of all nonzero training samples do not have to be the same. We just assume that the two parts are identical, for the sake of lower complexity. Within each part, the phases of training samples can be arbitrary. The P-BLUE-2 frequency-offset estimate can simply be obtained from the correlation between the two parts as

$$\hat{v} = \frac{N}{2\pi(N-K)} \angle \left\{ \sum_{k=0}^{K-1} r(k) r^*(k + N - K) \right\}. \qquad (30)$$

This estimator assumes that the normalized frequency offset v satisfies the condition $|v| < N/(2N - 2K)$. For the situations where v does not satisfy this condition, we propose the following approaches.

C. P-BLUE-3 Method

To handle a larger frequency offset, we modify the approximate version of the optimal training-signal structure to be composed of three nonzero parts: $\{s_0, \ldots, s_{K-1}\}$, $\{s_M, \ldots, s_{M+K-1}\}$, and $\{s_{N-K}, \ldots, s_{N-1}\}$ where $K = E_s/(3\tilde{P})$, $M \geq K$, and M should be chosen to satisfy the condition $|v| < N/(2M)$. The second part can also be set as $\{s_{N-M-K}, \ldots, s_{N-M-1}\}$, but in the following derivation, we use the former structure. For simplicity and

$$R(m) = \begin{cases} \displaystyle\sum_{k=0}^{K-1-m} r_k^* r_{k+m} + \sum_{k=N-K}^{N-1-m} r_k^* r_{k+m}, & 1 \leq m \leq K-1 \\ \displaystyle\sum_{k=N-K-m}^{K-1} r_k^* r_{k+m}, & N-2K+1 \leq m \leq N-K-1 \\ \displaystyle\sum_{k=0}^{N-1-m} r_k^* r_{k+m}, & N-K \leq m \leq N-1 \end{cases} \qquad (25)$$

$$R(m) = \begin{cases} \displaystyle\sum_{k=0}^{K-1-m} r_k^* r_{k+m} + \sum_{k=N-K}^{N-1-m} r_k^* r_{k+m}, & 1 \leq m \leq N-2K \\ \displaystyle\sum_{k=N-K-m}^{K-1} r_k^* r_{k+m}, & K \leq m < N-K \\ \displaystyle\sum_{k=0}^{N-1-m} r_k^* r_{k+m}, & N-K \leq m \leq N-1 \\ \displaystyle\sum_{k=0}^{K-1-m} r_k^* r_{k+m} + \sum_{k=N-K}^{N-1-m} r_k^* r_{k+m} + \sum_{k=N-K-m}^{K-1} r_k^* r_{k+m}, & N-2K+1 \leq m \leq K-1 \end{cases} \qquad (26)$$

lower complexity, the three nonzero parts are assumed to be identical. Then we have three estimates $\{\theta_k\}$ based on three part-wise correlation terms as

$$\theta_0 = \frac{N}{2\pi M} \angle \left\{ \sum_{k=0}^{K-1} r(k)r^*(k+M) \right\} \tag{31}$$

$$\beta_1 = \sum_{k=0}^{K-1} r(M+k)r^*(N-K+k) \tag{32}$$

$$\beta_2 = \sum_{k=0}^{K-1} r(k)r^*(N-K+k) \tag{33}$$

$$v_c = \frac{N}{2\pi M} \angle \{\beta_1^* \beta_2\} \tag{34}$$

$$\theta_1 = v_c + \frac{N}{2\pi(N-M-K)}$$
$$\times \angle \left\{ \beta_1 e^{-j(2\pi v_c(N-M-K)/N)} \right\} \tag{35}$$

$$\theta_2 = v_c + \frac{N}{2\pi(N-K)}$$
$$\times \angle \left\{ \beta_2 e^{-j(2\pi v_c(N-K)/N)} \right\} \tag{36}$$

where the use of v_c is to avoid a possible ambiguity in frequency-offset estimation based on correlation terms with larger correlation distances. The P-BLUE-3 frequency-offset estimate can then be obtained by (28) with $\boldsymbol{\theta} = [\theta_0, \theta_1, \theta_2]^T$ and $\boldsymbol{w}$ given by (29). Following Method B of [8], and after some calculation, we have the covariance matrix of $\{\theta_k\}$ given by

$$\boldsymbol{C_\theta} = \frac{3N}{8\pi^2\mathrm{SNR}}$$
$$\begin{bmatrix} \frac{2+\frac{3K}{N\cdot\mathrm{SNR}}}{M^2} & \frac{-1}{M(N-K-M)} & \frac{1}{M(N-K)} \\ \frac{-1}{M(N-K-M)} & \frac{2+\frac{3K}{N\cdot\mathrm{SNR}}}{(N-K-M)^2} & \frac{1}{(N-K-M)(N-K)} \\ \frac{1}{M(N-K)} & \frac{1}{(N-K-M)(N-K)} & \frac{2+\frac{3K}{N\cdot\mathrm{SNR}}}{(N-K)^2} \end{bmatrix}$$
$$\tag{37}$$

where $\mathrm{SNR} = E_s/(N\sigma_n^2)$. The required inverse covariance matrix is then given by equation (38) at the bottom of the page. Substituting it into (29) yields the BLUE weighting vector

$$\boldsymbol{w} = \frac{[M^2, (N-K-M)^2, (N-K)^2]^T}{(M^2 + (N-K-M)^2 + (N-K)^2)}. \tag{39}$$

Note that although $\boldsymbol{C_\theta}^{-1}$ depends on SNR values, $\boldsymbol{w}$ does not.

D. P-BLUE-4 Method

The result from Section III implies that allocating training energy on four samples is better than on three samples (see also [10]). This result may be applied to the modified training

signal by extending the sample-wise result to the part-wise result. For example, the modified training signal consisting of four parts may be better than that consisting of three parts. In order to investigate this, in the following, we derive a BLUE frequency-offset estimation method for the modified training signal consisting of four nonzero parts $\{s_0, \ldots, s_{K-1}\}$, $\{s_M, \ldots, s_{M+K-1}\}$, $\{s_{N-M-K}, \ldots, s_{N-M-1}\}$, and $\{s_{N-K}, \ldots, s_{N-1}\}$ where $M \geq K$ and M should be chosen to satisfy the condition $|v| < N/(2M)$. For simplicity, we assume that these four parts are phase-identical. To keep the same peak training-sample energy as in the three-part training signal, each sample of the first and fourth parts has energy $E_s/(3K)$, and that of the second and third parts has energy $E_s/(6K)$, where K is the same as in the three-part training signal.

Define the following:

$$R_0 = \sum_{k=0}^{K-1} \{r^*(k)r(k+M)$$
$$+ r^*(N-M-K+k)r(N-K+k)\} \tag{40}$$

$$R_1 = \sum_{k=0}^{K-1} r^*(M+k)r(N-M-K+k) \tag{41}$$

$$R_2 = \sum_{k=0}^{K-1} \{r^*(k)r(N-M-K+k)$$
$$+ r^*(M+k)r(N-K+k)\} \tag{42}$$

$$R_3 = \sum_{k=0}^{K-1} r^*(k)r(N-K+k). \tag{43}$$

Then we have four estimates $\{\theta_k : 0 \leq k \leq 3\}$ as follows:

$$\theta_0 = \frac{N}{2\pi M} \angle \{R_0\} \tag{44}$$

$$\theta_1 = \beta + \frac{N}{2\pi(N-2M-K)}$$
$$\times \angle \{R_1 e^{-j2\pi\beta(N-2M-K)/N}\} \tag{45}$$

$$\theta_2 = \beta + \frac{N}{2\pi(N-M-K)}$$
$$\times \angle \{R_2 e^{-j2\pi\beta(N-M-K)/N}\} \tag{46}$$

$$\theta_3 = \beta + \frac{N}{2\pi(N-K)} \angle \{R_3 e^{-j2\pi\beta(N-K)/N}\} \tag{47}$$

$$\beta = \frac{N}{2\pi M} \angle \{R_2^* R_3\}. \tag{48}$$

The P-BLUE-4 frequency-offset estimate can then be obtained by (28) with $\boldsymbol{\theta} = [\theta_0, \theta_1, \theta_2, \theta_3]^T$ and $\boldsymbol{w}$ given by (29). The covariance matrix required in the calculation of $\boldsymbol{w}$ is given by (49), shown at the bottom of the next page. Note that $\boldsymbol{C_\theta}$ and $\boldsymbol{w}$ are functions of SNR. However, it can be checked that $\boldsymbol{w}$ is not sensitive to SNR.

$$\boldsymbol{C_\theta}^{-1} = \frac{8\pi^2\mathrm{SNR}^2}{27K\left(1+\frac{K}{N\cdot\mathrm{SNR}}\right)} \begin{bmatrix} M^2(1+\frac{3K}{N\cdot\mathrm{SNR}}), & M(N-K-M), & -M(N-K) \\ M(N-K-M), & (N-K-M)^2(1+\frac{3K}{N\cdot\mathrm{SNR}}), & -(N-K-M)(N-K) \\ -M(N-K), & -(N-K-M)(N-K), & (N-K)^2(1+\frac{3K}{N\cdot\mathrm{SNR}}) \end{bmatrix} \tag{38}$$

V. COMPLEXITY, SIMULATION RESULTS, AND DISCUSSIONS

The total training-signal length is assumed to be $N = 1024$. The system can be either a single-carrier system or an OFDM system. To evaluate the proposed training signals, we obtain the CRBs analytically and the estimation MSE by simulation for the following four training signals. The first one, denoted by "Opt.," is the optimal training signal containing $2K$ nonzero samples with the peak training-sample energy P, and the second, denoted by "Ap. Opt.," is the approximate version of the optimal training signal having $2K$ nonzero samples with equal training-sample energy $\tilde{P} \leq P$. The third one, denoted by "Con-1," is a conventional training structure which has $2K$ nonzero samples with equal training-sample energy $\tilde{P}$, but all nonzero samples are consecutively located at the beginning of the training block. The last training signal, denoted by "Con-2," is also a conventional training structure which has N nonzero samples with equal energy of E_s/N. The estimation method used is MLE#1 from [13].

We can express the relation among E_s, P, and K as $E_s = P(2(K-1) + c)$, where $0 < c \leq 2$. If $c = 2$, Ap. Opt. becomes identical to Opt. In this evaluation, we assume without loss of generality that $c = 1/8$. We will use the peak-to-average sample energy ratio (PAR) of the training signal (which is a more common figure in practice) instead of the peak training-sample energy. PAR is defined as $\max\{|s_k|^2\}/(E_s/N)$ where $\{s_k\}$ represent training samples, including possibly zero-amplitude samples. The PARs for the above four training signals are given by $N/(2K - 15/8)$, $N/(2K)$, $N/(2K)$, and 1, respectively. Note that for a fixed E_s, depending on the allowable peak sample-energy constraint P or the allowable PAR constraint, the value of K will vary.

In Figs. 1–3, the CRBs and MSEs of the above four training signals are presented for different maximum allowable PAR values of 12.17, 9.10, and 6.06 dB, corresponding to the K values of 32, 64, and 128. The following remarks are in order.

1) The CRB advantage of the proposed structures over conventional ones is quite significant.
2) A larger peak-energy constraint (a larger PAR or a smaller K) gives a larger improvement to the optimal or approximate-optimal training signals.
3) The CRB or MSE performance difference between the optimal and the approximate-optimal training signals diminishes as K increases, and is negligible even at a small value of K. Hence, for the convenience in training-signal generation and deriving estimation method, the approximate version of the optimal training signal may be preferred.
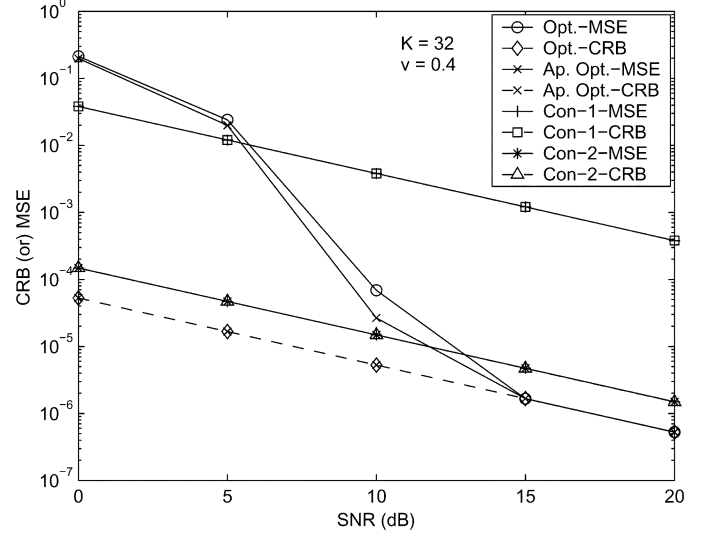

Fig. 1.   Comparison of CRB and MSE (obtained with MLE#1) between the proposed training structures and the conventional structures for $K = 32$.
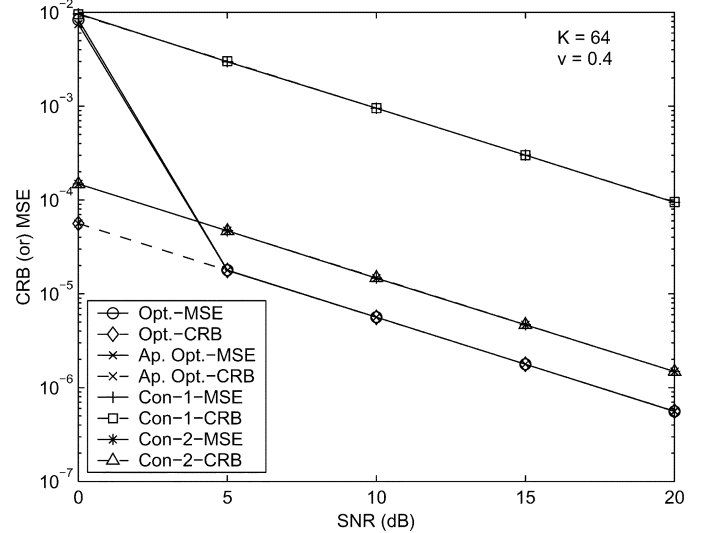

Fig. 2.   Comparison of CRB and MSE (obtained with MLE#1) between the proposed training structures and the conventional structures for $K = 64$.

4) For conventional structures, under the same total training-energy constraint, using more time resource as in "Con-2" gives a better performance than "Con-1." Note that the proposed structures use the same amount of time resource as "Con-1" does.
5) For the conventional structures, MLE#1 gives an MSE almost equal to CRB for all values of K and SNR. For

$$C_{\theta} = \frac{N}{4\pi^2 \mathrm{SNR}} \begin{bmatrix} \frac{9}{2}(\frac{1}{2} + \frac{K}{N\mathrm{SNR}}), & \frac{-3}{M(N-2M-K)}, & \frac{-3}{M(N-M-K)}, & \frac{3}{M(N-K)} \\ \frac{-3}{M(N-2M-K)}, & \frac{18(\frac{1}{3} + \frac{K}{N\mathrm{SNR}})}{(N-2M-K)^2}, & \frac{3}{(N-M-K)(N-2M-K)}, & 0 \\ \frac{-3}{M(N-M-K)}, & \frac{3}{(N-M-K)(N-2M-K)}, & \frac{\frac{9}{2}(\frac{1}{2} + \frac{K}{N\mathrm{SNR}})}{(N-M-K)^2}, & \frac{3}{(N-M-K)(N-K)} \\ \frac{3}{M(N-K)}, & 0, & \frac{3}{(N-M-K)(N-K)}, & \frac{\frac{9}{2}(\frac{2}{3} + \frac{K}{N\mathrm{SNR}})}{(N-K)^2} \end{bmatrix} \qquad (49)$$
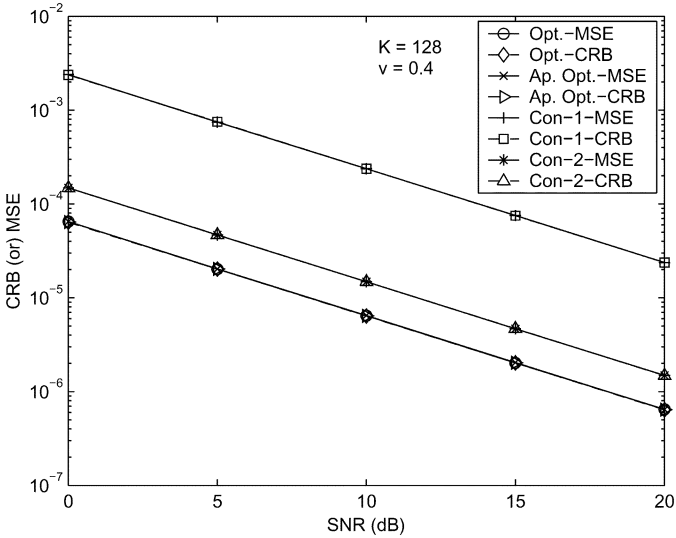
Fig. 3. Comparison of CRB and MSE (obtained with MLE#1) between the proposed training structures and the conventional structures for $K = 128$.
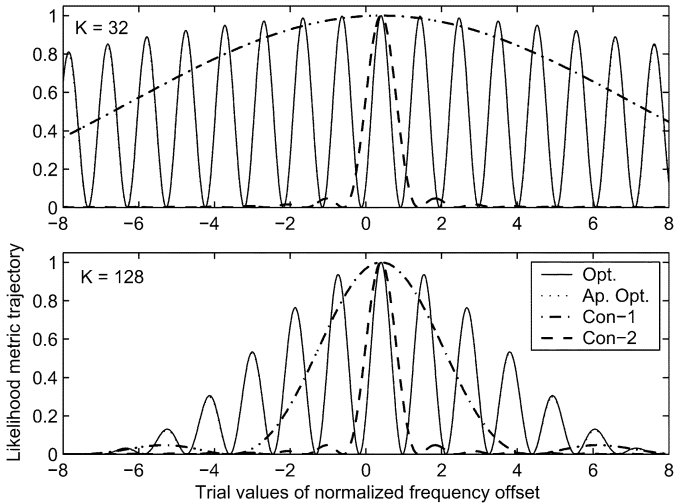

Fig. 4. Normalized likelihood metric trajectories of the MLE#1 for several training structures ($v = 0.4$). (Metric trajectories for Opt. and Ap. Opt. are not distinguishable in the plot.).

the proposed structures, MLE#1 gives an MSE almost equal to CRB when the number of nonzero training samples is large (i.e., large K), and it fails to provide a reliable estimate when the number of nonzero training samples is small (i.e., small K), especially at moderate and low SNR. (Note that the proposed BLUE methods do not have this drawback.)

The reason for 5) can be best explained by the normalized likelihood metric trajectories of MLE#1, shown in Fig. 4. The metric trajectories for the proposed structures (Opt. and Ap. Opt.) are not distinguishable in the plot. The CRB or MSE advantage of the proposed structures can be related to the corresponding sharper metric trajectory around the (maximum) peak. For a smaller K, the metric trajectories of the proposed structures contain several side lobes with peaks comparable to the maximum peak, which, in turn, results in a large estimation error at moderate or low SNR.
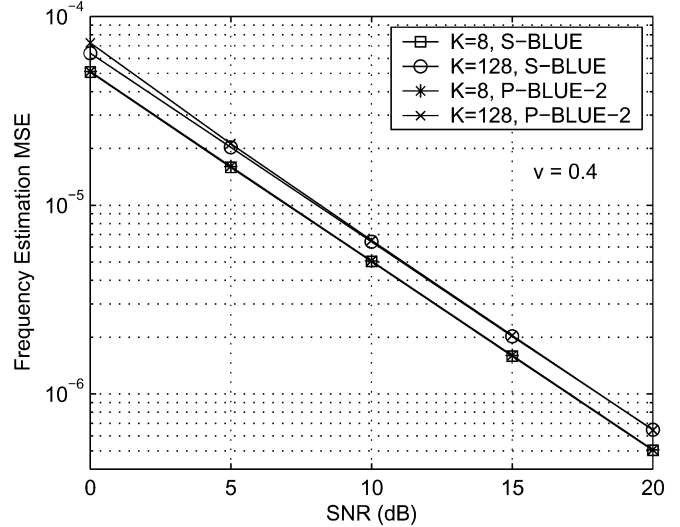

Fig. 5. Frequency-offset estimation MSE performance comparison between S-BLUE and P-BLUE-2 methods.

The frequency-offset estimation performance for the proposed S-BLUE and P-BLUE-2 methods using the Ap. Opt. structure is presented in Fig. 5. Both S-BLUE and P-BLUE-2 have essentially the same MSE performance (which are almost the same as the corresponding CRBs not shown in the figure for clarity). Both methods with $K = 8$(PAR = 18 dB) have better MSE performance than with $K = 128$(PAR = 6 dB), i.e., a larger allowable PAR gives a larger improvement. S-BLUE with $K = 8$ is just marginally better than P-BLUE-2 with $K = 8$. P-BLUE-2 has a larger estimation range, and a much lower complexity than S-BLUE (see Tables I and II). Hence, P-BLUE is more appealing than the S-BLUE method.

In Fig. 6, the performance comparison between P-BLUE-2 with the approximate-optimal training signal and the BLUE-based reference methods (Minn [9] and M&M [7]) with their corresponding training signal is presented. The reference training signal used for the reference methods is an OFDM training signal (a PN sequence in the frequency domain) consisting of $L = 8$ identical parts, as in [9]. The performance of P-BLUE-2 is evaluated for $K = 1, 8, 32, 64$, and 128, which correspond to the PAR values of 27, 18, 12, 9, and 6 dB, respectively. The reference training signal has a PAR value of about 6 dB.[1] The simulation results are in line with the theoretical results. The estimation performance is better for smaller values of K (i.e., a larger allowable PAR). The performances for $K = 1$ and $K = 128$ have approximately 1-dB SNR difference. The proposed approximate-optimal training signal with the proposed estimation method outperforms the reference training signal with the reference methods [7], [9]. The P-BLUE-2 method using the approximate-optimal training signal with $K = 1$ has about 5-dB SNR advantage over the reference methods using the reference training signal, while that with $K = 128$ has about 4-dB SNR advantage. The advantages are mainly due to the proposed training-signal structure, since all estimation methods used in Fig. 6 are based on the same BLUE principle.

[1]For an OFDM signal with 1024 subcarriers, the PAR value can occasionally be as high as 30 dB.

TABLE I
COMPUTATION LOADS OF FREQUENCY-OFFSET ESTIMATORS

Complexity		Real products	Real additions	$angle\cdot$
MLE#1	Opt. or Ap. Opt.	$4K[2 + Q'\ \eta'\ log_2(2Q'K)] + 2QN$	$2K[2 + 3Q'\ \eta'\ log_2(Q'K)] + QN$	
	Con-1	$4K[2 + Q'\eta'log_2(2Q'K)] + 2QN$	$2K[2 + 3Q'\eta'log_2(2Q'K)] + QN$	
	Con-2	$2N[2 + Q\ \eta\ log_2(QN)] + 2QN$	$N[2 + 3Q\ \eta\ log_2(QK)] + QN$	
S-BLUE		$8K^2 + 2K - 4$	$8K^2 - 7K + 1$	$3K - 2$
P-BLUE-2		$4K + 1$	$4K - 2$	1
P-BLUE-3		$12K + 19$	$12K + 4$	4
P-BLUE-4		$24K + 25$	$24K + 6$	5
M&M		$(2LK + 1)(L - 1)$	$2(L - 1)LK - 2$	$L - 1$
MINN		$2(LK + 1)(L - 1)$	$2(L - 1)LK - L$	$L - 1$

$$\eta = 1 - [log_2(Q) + 2(1/Q - 1)]/[log_2(QN)]$$

$$Q' = QN/(2K)$$

$$\eta' = 1 - [log_2(Q') + 2(1/Q' - 1)]/[log_2(Q'2K)]$$

TABLE II
A NUMERICAL EXAMPLE OF COMPUTATION LOADS OBTAINED WITH
$N = 1024$, $Q = 64$, $K = 64$ (OPT. OR AP. OPT.), $K = 43$ (MODIFIED
STRUCTURES), AND $L = 8$ (M&M OR MINN)

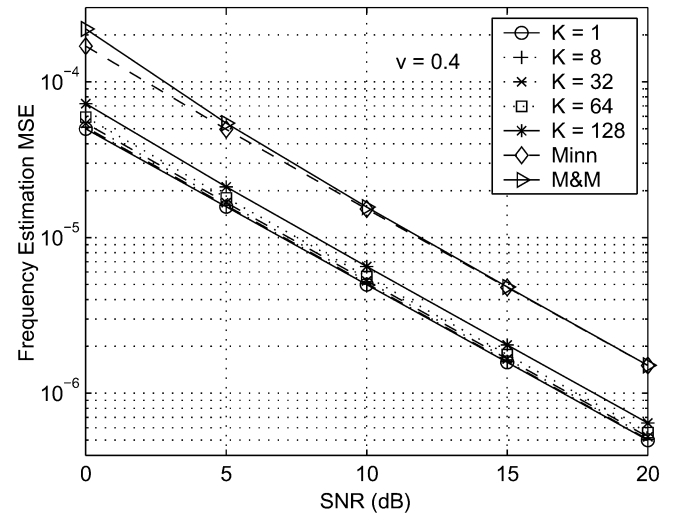| Complexity		Real products	Real additions	$angle\cdot$
MLE#1	Opt. or Ap. Opt.	1310720	1834496	
	Con-1	1310720	1834496	
	Con-2	1703936	2420736	
S-BLUE		32892	32321	190
P-BLUE-2		257	254	1
P-BLUE-3		535	520	4
P-BLUE-4		1057	1038	5
M&M		1799	1790	7
MINN		1806	1784	7

Fig. 6. Frequency-offset estimation MSE performance comparison between the proposed P-BLUE-2 method with the proposed training signal consisting of two nonzero parts, and the reference methods (Minn and M&M) using a training signal containing eight identical parts generated by a PN sequence in frequency domain.

Fig. 7 presents the performance comparison between the proposed P-BLUE-4 estimation method using the modified training signal with four nonzero parts, and the reference methods [7], [9] using the reference training signal. The performance is evaluated for $K = 1, 8, 32, 64$, and 128, which correspond to the PAR values of 25.33, 16.30, 10.28, 7.27, and 4.26 dB, respectively. The normalized frequency offset v is set to 1.6. $M = 128$ is used for all K values. The same observations as in Fig. 6 are obtained. A slightly different one from Fig. 6 is that the modified training signal (using P-BLUE-4 method) with $K = 1$ has

about 4 dB, and that with $K = 128$ has about 3-dB SNR advantage over the reference training signal (using the reference methods).

The performance comparison between the modified training signal with three nonzero parts and that with four nonzero parts, both having the same total training energy and the same PAR, is presented in Fig. 8. The P-BLUE-4 method using the modified training signal with four nonzero parts is slightly better than the P-BLUE-3 method using the modified training signal with three nonzero parts. This agrees with the implication of the result in
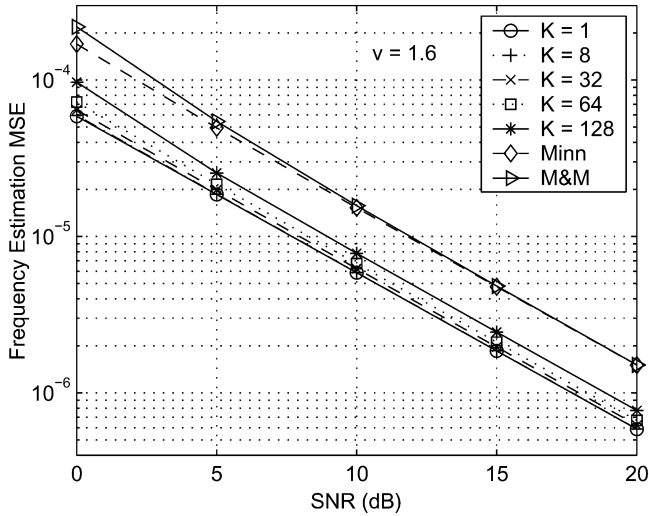
Fig. 7. Frequency-offset estimation MSE performance comparison between the proposed P-BLUE-4 method with the modified training signal consisting of four nonzero parts, and the reference methods (Minn and M&M) using a training signal containing eight identical parts generated by a PN sequence in frequency domain.
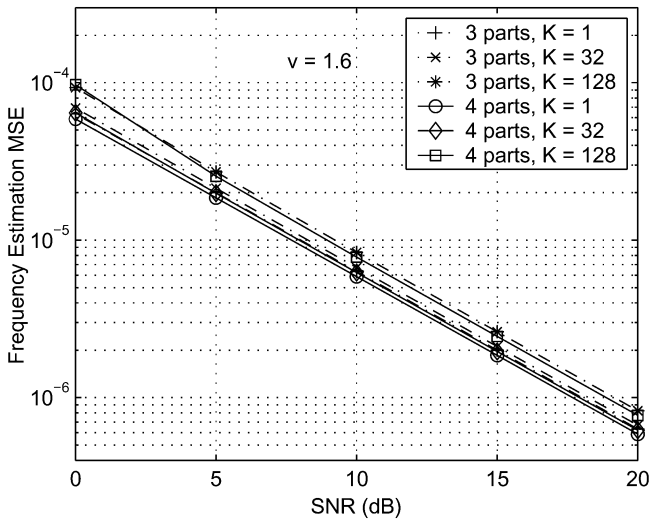

Fig. 8. Frequency-offset estimation MSE performance comparison between the proposed P-BLUE-3 method with the modified training signal consisting of three nonzero parts, and the proposed P-BLUE-4 method with the modified training signal consisting of four nonzero parts.

Section III. The slight performance improvement of the training signal with four nonzero parts over three nonzero parts is associated with a slightly larger complexity (see Tables I and II). We have also performed simulations for the reference methods using a training signal consisting of eight identical parts, where each part is a PN sequence in the time domain. The performance results (not shown) are almost the same as those obtained with the training signal consisting of eight identical parts generated by a PN sequence in the frequency domain. Note that the MSE performance of the proposed BLUE methods are almost the same as the corresponding CRBs, which are not shown in the figures for clarity.

The complexities of the estimation methods considered are presented in Table I, where MLE#1 uses a QN-point FFT and

a quadratic interpolation. The interpolation complexity is not included in the table. The training structure for S-BLUE and P-BLUE-2 is Ap. Opt., for P-BLUE-3 (or P-BLUE-4) is the modified structure with three (or four) nonzero parts, and for M&M and Minn is the reference OFDM training signal containing $L = 8$ consecutive identical parts. A numerical example for the complexities is given in Table II. To satisfy the same total training energy and the same peak-energy constraint, Table II uses $K = 64$ for MLE#1, S-BLUE, P-BLUE-2, and $K = 43$ for P-BLUE-3 and P-BLUE-4. The MLE#1 has a much higher complexity than the other methods. All P-BLUE methods have much smaller complexities than the others, and P-BLUE-2 has a smallest one. The smaller complexities of the proposed P-BLUE methods over the reference BLUE methods can be ascribed to the proposed training structures. We also observed in our simulation that the MLE#1 requires a larger fast Fourier transform (FFT) size (larger Q) for the training structures Opt. and Ap. Opt., and the Con-1 structure with a small K, to avoid an MSE floor at high SNR. The reason for Opt. and Ap. Opt. structures is the large side-lobe peaks of the likelihood metric, and that for Con-1 structure with a small K is a relatively flat likelihod metric around the maximum point (see Fig. 4).

An application of the proposed training-signal structures to multipath fading channels can be found in [12], where we constructed a training structure similar to the approximate-optimal training-signal structure for an OFDM-based wireless LAN in a multipath fading channel, and showed that it brings in a substantial performance improvement.

VI. CONCLUSIONS

An optimal training-signal structure for frequency-offset estimation with constraints on PAR, total training energy, and total training block length is developed based on minimizing the CRB. The proposed optimal training signal consists of two nonzero parts, where the length of each part is determined by the total training-signal energy and the allowable peak training-sample energy. For convenience in training-signal generation and estimator derivation, an approximate version of the optimal training-signal structure is proposed whose performance is practically not different from that of the optimal one. The existing ML method fails to provide a reliable estimate for the proposed optimal or approximate-optimal training signals with a small number of nonzero training samples at moderate or low SNR. Hence, we present two frequency-offset estimation methods (S-BLUE and P-BLUE-2) based on the BLUE principle which are robust to the parameter of the proposed training structures, and which have quite low complexity. To handle a larger frequency offset, a modified training signal consisting of three or four nonzero parts, and the corresponding part-wise correlation-based BLUE methods (P-BLUE-3 and P-BLUE-4), are presented. The proposed training signals and corresponding estimation methods achieve a substantial SNR advantage over the reference training signals using the reference estimation methods. The P-BLUE methods are more appealing than the S-BLUE method due to their lower complexity, larger estimation range, and essentially the same estimation MSE performance.

APPENDIX A
OPTIMALITY OF PROPOSED TRAINING-SIGNAL STRUCTURE

This section provides a proof for the optimality of the training-signal structure of (24). The optimization objective function (23) which is to be minimized can be given by the following objective function, which is to be maximized:

$$f(\boldsymbol{y}) = \boldsymbol{y}^T \boldsymbol{H} \boldsymbol{y} \quad \text{constraint: } \boldsymbol{y}^T \mathbf{1} = E_s \text{ and } 0 \leq y_i \leq P \quad (50)$$

where

$$\boldsymbol{H}_{k,n} = \begin{cases} -(G_{k,n} + G_{n,k}) = (k-n)^2, & k > n \\ 0, & \text{otherwise.} \end{cases} \quad (51)$$

We start from an arbitrary training signal. We apply on the training signal an algorithm which moves training-sample energy from one sample to another in a prescribed way. Every move will be shown to increase the objective function (50), which is to be maximized. The algorithm will finally reach to the training structure obtained in the previous section. Since transforming any training-signal structure into the training structure of (24) increases the objective function, the training structure of (24) is optimal and unique.

Let us consider a length-N training structure defined by

$$\boldsymbol{y} = [P \cdot \mathbf{1}_{1 \times M}, y_M, y_{M+1}, \dots, y_{N-L-1}, P \cdot \mathbf{1}_{1 \times L}]^T,$$
$$0 \leq M \leq N_1, \quad 0 \leq L \leq N_1, \quad M + L + 2 \leq N \quad (52)$$

where

$$0 \leq y_m \leq P, \quad 0 \leq m \leq N - 1 \quad (53)$$

$$\sum_{m=0}^{N-1} y_m = E_s, \quad \sum_{m=M}^{N-L-1} y_m = E_u, \quad N_1 = \left\lfloor \frac{E_s}{(2P)} \right\rfloor. \quad (54)$$

If there are more than N_1 consecutive samples of energy P at either side, the corresponding value of M or L is set to N_1. The above training structure represents any training signal satisfying the total energy and the peak energy constraints. The algorithm is described in the following. If $M \leq L$, then the algorithm moves energy ε from the training-energy sample y_k to y_M, where y_k with $M + 1 \leq k \leq N - L - 1$ is the nearest nonzero sample to y_M, and ε is given by (55), shown at the bottom of the page.

If $M \geq L$, the algorithm moves energy ε from the training-energy sample y_k to y_{N-L-1}, where y_k with $M \leq k \leq N - L - 2$ is the nearest nonzero sample to y_{N-L-1}, and ε is given by (56), shown at the bottom of the page.

Then the values of M and L are updated, and the energy moving is repeated. Finally, the training-energy structure will become that of (24). In the following, we will prove that each move increases the objective function which is to be maximized.

Each move changes the training structure $\boldsymbol{y} = [y_0, y_1, \dots, y_l, \dots, y_n, \dots, y_{N-1}]^T$ into $\boldsymbol{y}' = [y_0, y_1, \dots, y_l', \dots, y_n', \dots, y_{N-1}]^T$. The only differences between $\boldsymbol{y}$ and $\boldsymbol{y}'$ are at the indexes l and n. The objective function (50) can be given by (57), where Q is not a function of y_l and/or y_n

$$
\begin{aligned}
f(\boldsymbol{y}) = {} & y_l P \left[M l^2 + M(M-1)\left(\frac{2M-1}{6} - l\right) \right. \\
& \left. + L(L+1)\left(\frac{2L+1}{6} - N + l\right) \right. \\
& \left. + L(N-l)^2 \right] \\
& + y_n P \left[M n^2 + M(M-1)\left(\frac{2M-1}{6} - n\right) \right. \\
& \left. + L(L-1)\left(\frac{2L-1}{6} + N - L - n\right) \right. \\
& \left. + L(N-L-n)^2 \right] \\
& + y_l \left[\sum_{i=M}^{l-1} y_i (l-i)^2 + \sum_{i=l+1}^{n-1} y_i (l-i)^2 \right. \\
& \left. + \sum_{i=n+1}^{N-L-1} y_i (l-i)^2 \right] \\
& + y_n \left[\sum_{i=M}^{l-1} y_i (n-i)^2 + \sum_{i=l+1}^{n-1} y_i (n-i)^2 \right. \\
& \left. + \sum_{i=n+1}^{N-L-1} y_i (n-i)^2 \right] + y_l y_n (n-l)^2 + Q.
\end{aligned}
$$
$$(57)$$

Define $\Delta f = f(\boldsymbol{y}') - f(\boldsymbol{y})$. Then we have to prove that $\Delta f > 0$. Consider the case with $M \leq L$, where the energy ε is moved from y_k to y_M. Then we have $y_k = y_k - \varepsilon$ and $y_M = y_M + \varepsilon$. After substituting M and k in place of l and n in (57) and manipulating, we have

$$\Delta f = g_1 + g_2 \quad (58)$$

$$\varepsilon = \begin{cases} \min(P - y_M, y_k), & \text{if } (E_u \geq 2P) \text{ or } (E_u < 2P, M \neq L) \\ \min\left(\frac{E_u}{2} - y_M, y_k\right), & \text{if } \left(E_u < 2P, M = L, \text{ and } y_M < \frac{E_u}{2}\right) \\ 0, & \text{otherwise} \end{cases} \quad (55)$$

$$\varepsilon = \begin{cases} \min(P - y_{N-L-1}, y_k), & \text{if } (E_u \geq 2P) \text{ or } (E_u < 2P, M \neq L) \\ \min\left(\frac{E_u}{2} - y_{N-L-1}, y_k\right), & \text{if } E_u < 2P, M = L, \text{ and } y_{N-L-1} < \frac{E_u}{2} \\ 0, & \text{otherwise} \end{cases} \quad (56)$$

where g_1 and g_2 are, respectively, given by

$$
\begin{aligned}
g_1 &= \varepsilon P(k-M)L\left\{2N-k-M-L-1-(k+1)M/L\right\} \\
&= \left\{\varepsilon P(k-M)L(2N-2k-M-L-2)\right\} \\
&\quad + \left\{\varepsilon P(k-M)(k+1)(L-M)\right\} \\
&\geq \left\{\varepsilon P(k-M)(k+1)(L-M)\right\} \quad (59)
\end{aligned}
$$

$$
\begin{aligned}
g_2 &= \varepsilon(k-M)\left\{\sum_{i=k}^{N-L-1} y_i(k-M+2(i-k))\right\} \\
&\quad - \varepsilon(k-M)^2(y_M+\varepsilon) \\
&\geq \varepsilon(k-M)^2\left\{\left(\sum_{i=k}^{N-L-1} y_i\right)-(y_M+\varepsilon)\right\}. \quad (60)
\end{aligned}
$$

Then we obtain

$$
\begin{aligned}
\Delta f &\geq \left\{\varepsilon P(k-M)(k+1)(L-M)\right\} \\
&\quad + \varepsilon(k-M)^2\left\{\left(\sum_{i=k}^{N-L-1} y_i\right)-(y_M+\varepsilon)\right\}. \quad (61)
\end{aligned}
$$

When $(E_u > 2P)$ or $(E_u < 2P$ and $M = L)$, we have $(\sum_{i=k}^{N-L-1} y_i) > (y_M+\varepsilon)$, and the first term of (61) is greater than or equal to zero. Hence, $\Delta f > 0$. When $(E_u < 2P$ and $M < L)$, i.e., $(E_u < 2P$ and $M = L-1)$, we have $(\sum_{i=k}^{N-L-1} y_i)-(y_M+\varepsilon) > -P$, and the first term of (61) is $\varepsilon P(k-M)(k+1)$. Hence, we still have $\Delta f > 0$.

Now, consider the case with $M \geq L$, where the energy ε is moved from y_k to y_{N-L-1}. Then we have $y_k = y_k - \varepsilon$ and $y_{N-L-1} = y_{N-L-1}+\varepsilon$. After substituting k and $(N-L-1)$ in place of l and n in (57) and manipulating, we have (58) with g_1 and g_2, respectively, given by

$$
\begin{aligned}
g_1 &= \varepsilon P(N-L-k-1) \\
&\quad \times \left\{M(N-L-M+k)-L(N-k)\right\} \\
&= \varepsilon P(N-L-k-1) \\
&\quad \times \left\{M(2k-M-L)+(M-L)(N-k)\right\} \\
&\geq \varepsilon P(N-L-k-1)\left\{(M-L)(N-k)\right\} \quad (62)
\end{aligned}
$$

$$
\begin{aligned}
g_2 &= \varepsilon(N-L-k-1) \\
&\quad \times \left\{\sum_{i=M}^{k} y_i(N-L-k-1+2(k-i))\right\} \\
&\quad - \varepsilon(N-L-k-1)^2(y_{N-L-1}+\varepsilon) \\
&\geq \varepsilon(N-L-k-1)^2 \\
&\quad \times \left\{\left(\sum_{i=M}^{k} y_i\right)-(y_{N-L-1}+\varepsilon)\right\}. \quad (63)
\end{aligned}
$$

Consequently, we obtain

$$
\begin{aligned}
\Delta f &\geq \varepsilon P(N-L-k-1)\left\{(M-L)(N-k)\right\} \\
&\quad + \varepsilon(N-L-k-1)^2\left\{\left(\sum_{i=M}^{k} y_i\right)-(y_{N-L-1}+\varepsilon)\right\}. \quad (64)
\end{aligned}
$$

When $(E_u > 2P)$ or $(E_u < 2P$ and $M = L)$, we have $(\sum_{i=M}^{k} y_i) > (y_{N-L-1}+\varepsilon)$, and the first term of (64) is greater than or equal to zero. Hence, $\Delta f > 0$. When $(E_u < 2P$ and $M > L)$, i.e., $(E_u < 2P$ and $M = L+1)$, we have $\{(\sum_{i=M}^{k} y_i)-(y_{N-L-1}+\varepsilon)\} > -P$, and the first term of (61) is $\varepsilon P(N-L-k-1)(N-k)$. Hence, we still have $\Delta f > 0$.

The above proof assumes $M+L+2 \leq N$. The only possibility left is $M+L+1 = N$, which is already in the form of the training-signal structure (24). This proves that the training-signal (energy) structure (24) is optimal and unique.

APPENDIX B

This appendix provides the covariance matrix of $\boldsymbol{\theta}$ for S-BLUE method. Following Method B of [8] and skipping detailed calculation, we obtain the elements of the covariance matrix given as follows.

Case I: $N \geq 3K - 1$

If $1 \leq m \leq K-1, 1 \leq n \leq K-1$

$$
\begin{aligned}
C_\theta(m,n) &= \frac{N^2}{8\pi^2 mn(K-m)(K-n)\text{SNR}} \\
&\times \begin{cases} \frac{K-m}{2\text{SNR}}+m, & m=n \ \& \ m < \frac{K}{2} \\ \frac{K-m}{2\text{SNR}}+(K-m), & m=n \ \& \ m \geq \frac{K}{2} \\ \min(m,n), & m\neq n \ \& \ m+n < K \\ K-\max(m,n), & m\neq n \ \& \ m+n \geq K. \end{cases}
\end{aligned}
\quad (65)
$$

If $N-2K+1 \leq m \leq N-K-1, N-2K+1 \leq n \leq N-K-1$, we have

$$
\begin{aligned}
C_\theta(m,n) &= \frac{N^2}{4\pi^2 mn(2K+m-N)(2K+n-N)\text{SNR}} \\
&\times \begin{cases} \frac{2K+m-N}{2\text{SNR}}+(2K+m-N), & m=n \\ 2K+\min(m,n)-N, & m \neq n. \end{cases}
\end{aligned}
\quad (66)
$$

If $N-K \leq m \leq N-1, N-K \leq n \leq N-1$

$$
\begin{aligned}
C_\theta(m,n) &= \frac{N^2}{4\pi^2 mn(N-m)(N-n)\text{SNR}} \\
&\times \begin{cases} \frac{N-m}{2\text{SNR}}+(N-m), & m=n \\ N-\max(m,n), & m \neq n. \end{cases}
\end{aligned}
\quad (67)
$$

If $1 \leq m \leq K-1, N-2K+1 \leq n \leq N-K-1$, we have (68), shown at the bottom of the page. If $1 \leq m \leq K-1$, $N-K \leq n \leq N-1$, we have

$$
\begin{aligned}
C_\theta(m,n) &= \frac{N^2}{8\pi^2 mn(K-m)(N-n)\text{SNR}} \\
&\times \begin{cases} K-\max(K-m,N-n), & N-1-n \geq m \\ \min(K-m,N-n), & N-1-n < m. \end{cases}
\end{aligned}
\quad (69)
$$

$$
C_\theta(m,n) = \frac{-N^2}{8\pi^2 mn(K-m)(2K+n-N)\text{SNR}} \times \begin{cases} \min(m,N-K-n), & N-K-n \leq K-1-m \\ K-\max(m,N-K-n), & N-K-n > K-1-m \end{cases} \quad (68)
$$

If $1 \leq m \leq K - 1$, $N - K \leq n \leq N - 1$

$$
C_\theta(m,n) = \frac{N^2}{4\pi^2 mn(2K+m-N)(N-n)\text{SNR}}
$$
$$
\times \begin{cases} K-n+m, & K-1+m \geq n \\ 0, & K-1+m < n. \end{cases} \quad (70)
$$

Case II: $N \leq 3K - 2$
If $1 \leq m \leq N - 2K$, $1 \leq n \leq N - 2K$, we have

$$
C_\theta(m,n) = \frac{N^2}{8\pi^2 mn(K-m)(K-n)\text{SNR}}
$$
$$
\times \begin{cases} \frac{K-m}{2\text{SNR}} + m, & m = n \ \& \ m < \frac{K}{2} \\ \frac{K-m}{2\text{SNR}} + (K-m), & m = n \ \& \ m \geq \frac{K}{2} \\ \min(m,n), & m \neq n \ \& \ m+n < K \\ K - \max(m,n), & m \neq n \ \& \ m+n \geq K. \end{cases}
$$
$$
(71)
$$

If $K \leq m \leq N - K - 1$, $K \leq n \leq N - K - 1$

$$
C_\theta(m,n) = \frac{N^2}{4\pi^2 mn(2K+m-N)(2K+n-N)\text{SNR}}
$$
$$
\times \begin{cases} \frac{2K+m-N}{2\text{SNR}} + (2K+m-N), & m = n \\ 2K + \min(m,n) - N, & m \neq n. \end{cases} \quad (72)
$$

If $N - K \leq m \leq N - 1$, $N - K \leq n \leq N - 1$

$$
C_\theta(m,n) = \frac{N^2}{4\pi^2 mn(N-m)(N-n)\text{SNR}}
$$
$$
\times \begin{cases} \frac{N-m}{2\text{SNR}} + (N-m), & m = n \\ N - \max(m,n), & m \neq n. \end{cases} \quad (73)
$$

If $1 \leq m \leq N - 2K$, $K \leq n \leq N - K - 1$, we have (74), shown at the bottom of the page. If $1 \leq m \leq N - 2K$, $N - K \leq n \leq N - 1$, we have

$$
C_\theta(m,n) = \frac{N^2}{8\pi^2 mn(K-m)(N-n)\text{SNR}}
$$
$$
\times \begin{cases} K - \max(K-m, N-n), & N-1-n \geq m \\ \min(K-m, N-n), & N-1-n < m. \end{cases}
$$
$$
(75)
$$

If $K \leq m \leq N - K$, $N - K \leq n \leq N - 1$

$$
C_\theta(m,n) = \frac{N^2}{4\pi^2 mn(2K+m-N)(N-n)\text{SNR}}
$$
$$
\times \begin{cases} K-n+m, & K-1+m \geq n \\ 0, & K-1+m < n. \end{cases} \quad (76)
$$

If $N - 2K + 1 \leq m \leq K - 1$, $N - 2K + 1 \leq n \leq K - 1$, we have (77), shown at the bottom of the page. If $1 \leq m \leq N - 2K$, $N - 2K + 1 \leq n \leq K - 1$, we have (78), shown at the bottom of the page. If $N - 2K + 1 \leq m \leq K - 1$,

---

$$
C_\theta(m,n) = \frac{-N^2}{8\pi^2 mn(K-m)(2K+n-N)\text{SNR}} \times \begin{cases} \min(m, N-K-n), & N-K-n \leq K-1-m \\ K - \max(m, N-K-n), & N-K-n > K-1-m \end{cases} \quad (74)
$$

$$
C_\theta(m,n) = \frac{N^2}{2\pi^2 mn(4K-m-N)(4K-n-N)\text{SNR}} \times \begin{cases} \frac{K-m}{2\text{SNR}} + m, & m = n \ \& \ m < \frac{K}{2} \\ \frac{K-m}{2\text{SNR}} + (K-m), & m = n \ \& \ m \geq \frac{K}{2} \\ \min(m,n), & m \neq n \ \& \ m+n < K \\ K - \max(m,n), & m \neq n \ \& \ m+n \geq K \end{cases}
$$
$$
+ \frac{-N^2}{4\pi^2 mn(4K-m-N)(4K-n-N)\text{SNR}} \times \begin{cases} \min(m, N-K-n), & N-K-n \leq K-1-m \\ K - \max(m, N-K-n), & N-K-n > K-1-m \end{cases}
$$
$$
+ \frac{N^2}{4\pi^2 mn(4K-m-N)(4K-n-N)\text{SNR}} \times \begin{cases} \frac{2K+m-N}{2\text{SNR}} + (2K+m-N), & m = n \\ 2K + \min(m,n) - N, & m \neq n \end{cases}
$$
$$
+ \frac{-N^2}{4\pi^2 mn(4K-m-N)(4K-n-N)\text{SNR}} \times \begin{cases} \min(m, N-K-n), & N-K-n \leq K-1-m \\ K - \max(m, N-K-n), & N-K-n > K-1-m \end{cases}
$$
$$
(77)
$$

$$
C_\theta(m,n) = \frac{N^2}{4\pi^2 mn(K-m)(4K-n-N)\text{SNR}} \times \begin{cases} \frac{K-m}{2\text{SNR}} + m, & m = n \ \& \ m < \frac{K}{2} \\ \frac{K-m}{2\text{SNR}} + (K-m), & m = n \ \& \ m \geq \frac{K}{2} \\ \min(m,n), & m \neq n \ \& \ m+n < K \\ K - \max(m,n), & m \neq n \ \& \ m+n \geq K \end{cases}
$$
$$
+ \frac{-N^2}{8\pi^2 mn(K-m)(4K-n-N)\text{SNR}} \times \begin{cases} \min(m, N-K-n), & N-K-n \leq K-1-m \\ K - \max(m, N-K-n), & N-K-n > K-1-m \end{cases}
$$
$$
(78)
$$

$$C_\theta(m,n) = \frac{-N^2}{4\pi^2 mn(4K-m-N)(2K+n-N)\text{SNR}} \times \begin{cases} \min(m, N-K-n), & N-K-n \le K-1-m \\ K-\max(m, N-K-n), & N-K-n > K-1-m \end{cases}$$
$$+ \frac{N^2}{4\pi^2 mn(4K-m-N)(4K-n-N)\text{SNR}} \times \begin{cases} \frac{2K+m-N}{2\text{SNR}} + (2K+m-N), & m=n \\ 2K+\min(m,n)-N, & m \ne n \end{cases} \tag{79}$$

$K \le n \le N-K-1$, we have (79), shown at the top of the page. If $N-2K+1 \le m \le K-1$, $N-K \le n \le N-1$, we have

$$C_\theta(m,n) = \frac{N^2}{4\pi^2 mn(4K-m-N)(N-n)\text{SNR}}$$
$$\times \begin{cases} K-\max(K-m, N-n), & N-1-n \ge m \\ \min(K-m, N-n), & N-1-n < m \end{cases}$$
$$+ \frac{N^2}{4\pi^2 mn(4K-m-N)(N-n)\text{SNR}}$$
$$\times \begin{cases} K-n+m, & K-1+m \ge n \\ 0, & K-1+m < n. \end{cases} \tag{80}$$

ACKNOWLEDGMENT

The authors thank the reviewers and the Editor for their valuable comments.

REFERENCES

[1] 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, pp. 230–238, Feb. 1999.

[2] W. Chen and U. Mitra, "Training sequence optimization: Comparison and an alternative criterion," *IEEE Trans. Commun.*, vol. 48, pp. 1987–1991, Dec. 2000.

[3] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEE Trans. Veh. Technol.*, vol. 49, pp. 1207–1215, Jul. 2000.

[4] H. Minn, M. Zeng, and V. K. Bhargava, "On timing offset estimation for OFDM systems," *IEEE Commun. Lett.*, vol. 4, pp. 242–244, Jul. 2000.

[5] H. Minn, V. K. Bhargava, and K. B. Letaief, "A robust timing and frequency synchronization for OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 2, pp. 822–839, Jul. 2003.

[6] T. M. Schmidl and D. C. Cox, "Robust frequency and timing synchronization for OFDM," *IEEE Trans. Commun.*, vol. 45, pp. 1613–1621, Dec. 1997.

[7] M. Morelli and U. Mengali, "An improved frequency offset estimator for OFDM applications," *IEEE Commun. Lett.*, vol. 3, pp. 75–77, Mar. 1999.

[8] 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.

[9] ——, "Some issues of complexity and training symbol design for OFDM frequency offset estimation methods based on BLUE principle," in *Proc. IEEE Veh. Technol. Conf.*, Seoul, Korea, Apr. 2003, pp. 1288–1292.

[10] H. Minn, "An investigation into training signal design for frequency offset estimation," in *Proc. 41st Annu. Allerton Conf. Commun., Control, Comput.*, Urbana-Champaign, IL, Oct. 2003, pp. 1735–1744.

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

[12] H. Minn and V. K. Bhargava, "A reduced complexity, improved frequency offset estimation for OFDM-based WLAN's," in *Proc. WNCG Wireless Networking Symp.*, Austin, TX, Oct. 2003.

[13] M. Morelli and U. Mengali, "Carrier-frequency estimation for transmissions over selective channels," *IEEE Trans. Commun.*, vol. 48, pp. 1580–1589, Sep. 2000.

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 the Asian Institute of Technology (AIT), Bangkok, Thailand, in 1997, and 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, University of Victoria. Since September 2002, he has been with the Erik Jonsson School of Engineering and Computer Science, University of Texas at Dallas, Richardson, TX, as an Assistant Professor. His research interests include wireless communications, statistical signal processing, error control, detection, estimation, synchronization, signal design, and cross-layer design.
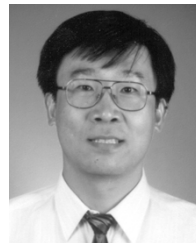
Shaohui Xing was born in Hebei, China, in 1970. He received the B.S. and M.S. degrees from the National University of Defense Technology, Changsha, China, in 1993 and 1996, respectively, and the M.S. degree in telecommunication engineering from University of Texas at Dallas, Richardson, TX, in 2002, where he is currently working toward the Ph.D. degree in the Department of Electrical Engineering.

His current research interests focus on multicarrier communications; specifically, channel estimation, and timing and frequency synchronization for OFDM systems.