Research Article
Constrained Optimization of MIMO Training Sequences

Justin P. Coon and Magnus Sandell

Toshiba Telecommunications Research Laboratory, 32 Queen Square, Bristol BS1 4ND, UK

Multiple-input multiple-output (MIMO) systems have shown a huge potential for increased spectral efficiency and throughput. With an increasing number of transmitting antennas comes the burden of providing training for channel estimation for coherent detection. In some special cases optimal, in the sense of mean-squared error (MSE), training sequences have been designed. However, in many practical systems it is not feasible to analytically find optimal solutions and numerical techniques must be used. In this paper, two systems (unique word (UW) single carrier and OFDM with nulled subcarriers) are considered and a method of designing near-optimal training sequences using nonlinear optimization techniques is proposed. In particular, interior-point (IP) algorithms such as the barrier method are discussed. Although the two systems seem unrelated, the cost function, which is the MSE of the channel estimate, is shown to be effectively the same for each scenario. Also, additional constraints, such as peak-to-average power ratio (PAPR), are considered and shown to be easily included in the optimization process. Numerical examples illustrate the effectiveness of the designed training sequences, both in terms of MSE and bit-error rate (BER).

1. INTRODUCTION

Future wireless systems can offer substantially higher data rates than current systems by using new, sophisticated technologies. One of the most promising technologies is multiple-input multiple-output (MIMO) transmission [1, 2], where spatial multiplexing [3, 4], or more advanced space-time codes [5–7], can increase the spectral efficiency by using the spatial domain. One drawback with MIMO systems is that channel estimation becomes more important. Not only are MIMO decoders more sensitive to channel estimation errors than their single-antenna counterparts, the overhead in terms of required training sequences is also increased. Thus, it is important to make training as efficient as possible.

For a MIMO system with M transmit antennas, the simplest form of training sequence is to transmit from only one antenna at a time. This method, however, requires M slots, which in some cases can be a large overhead. One technique that has been applied in MIMO orthogonal frequency division multiplexing (OFDM) systems (see, e.g., [8] for an overview of OFDM) to reduce this overhead is to exploit the channel dimensions. Since OFDM systems design the data in the frequency domain, channel estimates are required for all K subcarriers. However, the time-domain channel impulse response (CIR) is often assumed to be shorter than the length-Q cyclic prefix (CP), where typically $Q \ll K$. Hence, the frequency-domain channel lies in a subspace, which makes it possible to transmit training sequences simultaneously from several antennas (in fact, K/Q) [9–12]. Similar techniques have been shown to work for single-carrier MIMO systems [13].

In general, it is not enough to simply transmit training sequences simultaneously from several antennas in a MIMO system. Indeed, the quality of the channel estimate is, in many cases, just as important as obtaining an estimate efficiently. Designing training sequences such that they facilitate high-quality MIMO channel estimation is a topic that has seen much research for OFDM and single-carrier systems alike (see, e.g., [10, 12–16]). Sequences that minimize (or maximize) some cost function associated with the quality of the channel estimate are said to be *optimal*. In many ideal scenarios, training sequences possessing optimal properties for MIMO channel estimation can be designed analytically. A typical metric that is used to measure the quality of a channel estimate, and thus the optimality of training sequences, is the mean-squared error (MSE) of the channel estimate. Sequences that minimize the MSE of a MIMO channel estimate

have been designed analytically for OFDM systems [10, 12] as well as single-carrier systems with a CP extension [13].

In addition to providing optimal channel estimation, it is often desirable for MIMO training sequences to possess other benefits that facilitate implementation of the sequences. One such benefit that is commonly required of training sequences is that they have a low peak-to-average power ratio (PAPR). An example of optimal MIMO training sequences that have good PAPR properties can be found in [13].

One practical point that many researchers have overlooked while designing training sequences for MIMO OFDM is the fact that many OFDM systems require some subcarriers to be nulled. Nulled subcarriers are typically used for spectral shaping to ensure that the OFDM signal fits within a given spectral mask. Similarly, some single-carrier systems utilize a unique word (UW) (a.k.a. known symbol padding (KSP)) to estimate the channel [17, 18]. These known sequences can be viewed as training sequences with nulled symbols that are superimposed onto data sequences prior to transmission. When a portion of a training sequence comprises nulled symbols, as in the two previous examples, the quality of the channel estimate often suffers. For the example of MIMO OFDM, it was pointed out in [19] that training sequence designs that are optimal when all subcarriers are used no longer achieve the lower bound on MSE when nulled subcarriers are employed.

Unfortunately, once constraints such as nulled symbols are introduced, the problem of optimal training sequence design often becomes analytically unsolvable. In many of these cases, one may turn to numerical methods to find optimal training sequences, such as those proposed in [20, 21]. If, however, additional constraints are added to the training sequences, such as constraints on the PAPR of the sequences, more sophisticated nonlinear optimization techniques must be applied. Interior-point (IP) methods, for example, allow both equality and inequality constraints to be added to conventional optimization problems, which can then be solved in an efficient manner [22]. These methods have been applied to solve optimization problems in several areas, including power systems, network optimization, and MIMO transceiver design [22–25].

In this paper, IP methods are used to design MIMO training sequences under difficult constraints, such as nulled symbols and an upper limit on PAPR. Two specific scenarios are considered in order to demonstrate the efficacy of IP methods in the context of sequence design: MIMO OFDM with nulled subcarriers and single-carrier MIMO with a UW extension. In both of these scenarios, a least-squares (LS) MIMO channel estimator is considered, and it will be shown that sequences with near-optimal properties (in the MSE sense) can be found. It should be noted that the techniques proposed in this paper can be adapted for use with other estimators, such as the minimum mean-square error (MMSE) estimator; however, these estimators typically require additional knowledge about the MIMO channel, such as the covariance and power delay profile. The LS estimator is considered in this paper for its simplicity.

In Section 2, the two aforementioned scenarios are described, and the LS channel estimator is detailed for each system. The proposed approach for designing optimal sequences for the two example scenarios is discussed in Section 3. In Section 4, results are given in the form of MSE and error-rate curves for systems that employ training sequences obtained through the application of IP methods. Finally, conclusions are drawn in Section 5.

2. LEAST-SQUARES CHANNEL ESTIMATION IN MIMO SYSTEMS

Channel estimation in MIMO systems has received much attention in recent years (see, e.g., [10–12, 14–16]). A popular method of performing MIMO channel estimation is the LS method. LS channel estimation, which can be shown to be equivalent to maximum likelihood (ML) channel estimation when the noise in the system is white and Gaussian distributed [19], is simple to derive and implement, and can be generalized to many MIMO scenarios. In this section, the LS channel estimator is derived both for single-carrier MIMO systems using a UW extension and for MIMO OFDM systems with nulled subcarriers. Furthermore, an expression for the MSE of the LS estimator will be detailed for each example system. These expressions can be used with nonlinear optimization techniques, such as IP methods, to find optimal training sequences in the MSE sense.

2.1. MIMO unique word

The concept of using the UW in single-carrier block transmissions as an alternative to the well-known CP extension was presented in [26]. A UW is simply a short sequence of symbols that is appended to each data block in a single-carrier block transmission system. The UW remains constant from block to block, thus giving the illusion that the transmission is periodic in a similar manner to a CP extension, but without the need for postprocessing at the receiver. Using this block transmission structure facilitates the use of low-complexity frequency-domain equalization techniques at the receiver.

The constant nature of the UW is its key advantage, and several uses for the UW extension that utilize this property have been proposed, including synchronization, phase tracking, and channel estimation and tracking [17, 18, 27–31]. The typical MIMO system with a UW extension can be described as follows. Let the ith length-K block of symbols at the mth transmit antenna be denoted by $\mathbf{x}_m(i)$. This vector can be partitioned into a length-P vector $\mathbf{s}_m(i)$ of data symbols and a length-Q vector representing the UW, which is the same from block to block. An illustration of this block structure is depicted in Figure 1. In order to mitigate inter-block interference (IBI), it is assumed that $Q \geq L - 1$ where L is the length of the CIR. This condition also induces circularity in the system when the channel remains static for at least one block duration, which allows the ith
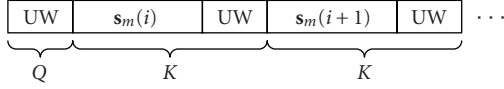
FIGURE 1: Example of UW block structure for single-carrier systems.

length-K block of symbols received at antenna n to be expressed by

$$\mathbf{y}_n(i) = \sum_{m=1}^{M} \mathbf{G}_{n,m}(i)\mathbf{x}_m(i) + \mathbf{v}_n(i), \qquad (1)$$

where M is the total number of transmit antennas, $\mathbf{G}_{n,m}(i)$ is a $K \times K$ circulant matrix representing the channel between the mth transmit antenna and the nth receive antenna at time i, and $\mathbf{v}_n(i)$ is a length-K vector of uncorrelated, zero-mean, complex Gaussian noise samples, each with a variance of $\sigma_v^2/2$ per dimension. The first column of $\mathbf{G}_{n,m}(i)$ is given by $(g_{n,m}(i,0),\ldots,g_{n,m}(i,L-1),0,\ldots,0)^T$, where $g_{n,m}(i,\ell)$ denotes the CIR coefficient for the (n,m)th channel at time i.

The circulant nature of the channel matrix facilitates frequency-domain processing of the received signal since diagonalization of the channel matrix is performed by pre- and postmultiplying the channel matrix by the $K \times K$ normalized discrete Fourier transform (DFT) and inverse DFT (IDFT) matrices, respectively. In other words, the matrix $\mathbf{H}_{n,m}(i) = \mathbf{F}\mathbf{G}_{n,m}(i)\mathbf{F}^H$ is a diagonal matrix where $h_{n,m}(i,k) = \sum_{\ell=0}^{L-1} g_{n,m}(i,\ell)\exp(-j2\pi k\ell/K)$ is the kth element of the diagonal and the (i,k)th element of the DFT matrix $\mathbf{F}$ is $F_{i,k} = 1/\sqrt{K}\exp(-j2\pi ik/K)$. A block diagram of a MIMO UW system that performs frequency-domain equalization on the received message is illustrated in Figure 2. Taking the DFT of the received vector $\mathbf{y}_n(i)$ gives

$$\tilde{\mathbf{y}}_n(i) = \sum_{m=1}^{M} \mathbf{H}_{n,m}(i)\mathbf{F}\mathbf{x}_m(i) + \tilde{\mathbf{v}}_n(i)$$
$$= \sum_{m=1}^{M} \mathbf{H}_{n,m}(i)\big(\mathbf{F}_P\mathbf{s}_m(i) + \mathbf{F}'_Q\mathbf{u}_m\big) + \tilde{\mathbf{v}}_n(i), \qquad (2)$$

where $\tilde{\mathbf{v}}_n(i) = \mathbf{F}\mathbf{v}_n(i)$, $\mathbf{u}_m$ is the UW for the mth transmit antenna, $\mathbf{F}_P$ denotes the first P columns of $\mathbf{F}$, and $\mathbf{F}'_Q$ denotes the last Q columns of $\mathbf{F}$. Assuming the channel remains static over, for example, B block durations and the transmitted data has a mean of zero, the data portion of the received message can be somewhat removed by averaging the corresponding B received vectors. This averaging can be expressed by

$$\overline{\mathbf{y}}_n = \frac{1}{B}\sum_{i=1}^{B} \tilde{\mathbf{y}}_n(i) = \sum_{m=1}^{M} \mathbf{H}_{n,m}\mathbf{F}'_Q\mathbf{u}_m + \tilde{\boldsymbol{\nu}}_n, \qquad (3)$$

where $\mathbf{H}_{n,m}(i) \equiv \mathbf{H}_{n,m}$ due to the static channel assumption and

$$\tilde{\boldsymbol{\nu}}_n = \frac{1}{B}\sum_{i=1}^{B}\sum_{m=1}^{M} \mathbf{H}_{n,m}\mathbf{F}_P\mathbf{s}_m(i) + \frac{1}{B}\sum_{i=1}^{B} \tilde{\mathbf{v}}_n(i). \qquad (4)$$

Note that since the data and noise are zero-mean and uncorrelated,

$$\lim_{B\to\infty} \tilde{\boldsymbol{\nu}}_n = E\{\tilde{\boldsymbol{\nu}}_n\} = \mathbf{0}_K, \qquad (5)$$

where $\mathbf{0}_K$ denotes the length-K column vector of zeros. Furthermore, the covariance matrix of $\tilde{\boldsymbol{\nu}}_n$ is assumed to be given by[1]

$$E\{\tilde{\boldsymbol{\nu}}_n\tilde{\boldsymbol{\nu}}_n^H\} = \tilde{\sigma}^2 \mathbf{I}_K, \qquad (6)$$

where $\tilde{\sigma}^2$ is the variance of each element of $\tilde{\boldsymbol{\nu}}_n$.

The stochastic model presented in (3) can be used to perform channel estimation in MIMO systems that use a UW extension [29]. Following the method of [10], (3) can be rewritten as

$$\overline{\mathbf{y}}_n = \sum_{m=1}^{M} \sqrt{K}\tilde{\mathbf{U}}_m\mathbf{F}_L\mathbf{g}_{n,m} + \tilde{\boldsymbol{\nu}}_n = \mathbf{A}\mathbf{g}_n + \tilde{\boldsymbol{\nu}}_n, \qquad (7)$$

where $\tilde{\mathbf{U}}_m := \mathrm{diag}\{\mathbf{F}'_Q\mathbf{u}_m\}$, $\mathbf{g}_{n,m}$ is a length-L vector composed of the CIR coefficients for the (n,m)th channel, $\mathbf{A} := \sqrt{K}(\tilde{\mathbf{U}}_1\mathbf{F}_L,\ldots,\tilde{\mathbf{U}}_M\mathbf{F}_L)$, and $\mathbf{g}_n := (\mathbf{g}_{n,1}^T,\ldots,\mathbf{g}_{n,M}^T)^T$. It is assumed that $K \geq ML$; thus, the matrix $\mathbf{A}$ is tall and has full column rank. Under this necessary condition, it follows that the LS channel estimate is given by

$$\hat{\mathbf{g}}_n = (\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\overline{\mathbf{y}}_n. \qquad (8)$$

From this expression, it is obvious that the channels can be estimated for each receive antenna separately; consequently, the index n is omitted from subsequent derivations and discussion. It should be noted that other channel estimators exist that outperform this first-order channel estimator; however, the emphasis here is on simplicity and the ability to design optimal (or nearly optimal) UWs for this practical estimator.

Typically, single-carrier systems employing a UW extension exploit the frequency domain to perform channel equalization [27, 28]. Thus, the frequency domain estimate of the channel is generally of more interest than the estimate of the CIR given above. This estimate is given by

$$\hat{\mathbf{h}} = (\mathbf{I}_M \otimes \mathbf{F}_L)\hat{\mathbf{g}}, \qquad (9)$$

where $\otimes$ denotes the Kronecker product operation. The MSE of this LS channel estimate is given by [10, 19]

$$\mathrm{MSE} = E\{\|\hat{\mathbf{h}} - \mathbf{h}\|^2\}$$
$$= \tilde{\sigma}^2 \mathrm{Tr}\{(\mathbf{I}_M \otimes \mathbf{F}_L)(\mathbf{A}^H\mathbf{A})^{-1}(\mathbf{I}_M \otimes \mathbf{F}_L)^H\}, \qquad (10)$$

where $\mathrm{Tr}\{\cdot\}$ denotes the trace operation.

In the limiting case where only one transmit antenna is used, it can be shown that the MSE term is minimized when

---

[1] Although the noise $\tilde{\boldsymbol{\nu}}_n$ is not strictly white due to the data term, this assumption facilitates the formulation of the LS channel estimator as shown below.
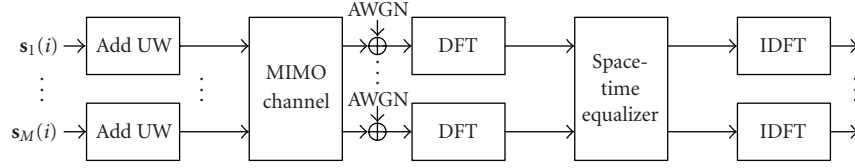
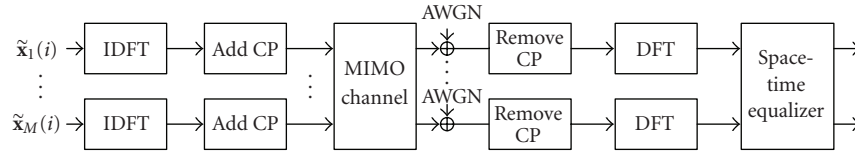FIGURE 2: Block diagram of a MIMO UW system that performs channel equalization in the frequency domain.

FIGURE 3: Block diagram of a MIMO OFDM system.

the partial DFT of the UW, given by $\mathbf{F}'_Q\mathbf{u}_1$, is constant modulus [32]. This result is intuitively satisfying since it implies that the channel frequency response coefficient for each frequency tone is given equal importance by the channel estimator. This observation extends to the MIMO case where the MSE of the channel estimate is minimized when $\mathbf{A}$ is a unitary matrix, which qualitatively implies that the DFT of each UW should have a constant modulus, but all UWs should be phase-shift orthogonal to each other [10]. When these conditions are satisfied, the channel between a given transmitter and the receiver is estimated optimally, as in the single-antenna case, and the signals from each transmit antenna are separable at the receiver, thus facilitating MIMO channel estimation. Unfortunately, UWs that have the properties described above do not exist in general [33]; however, nonlinear optimization techniques can be employed to find sequences that come arbitrarily close to providing optimal MIMO channel estimation in the MSE sense. These techniques will be discussed in Section 3.

One final note concerning the applicability of the UW in general MIMO systems should be made. By observing (1) and regarding the transmitted signal vector $\mathbf{x}_m(i)$ as comprising only the UW for the mth transmit antenna, that is, the data is perfectly removed—it is obvious that the mth signal vector $\mathbf{G}_{n,m}(i)\mathbf{x}_m$ has only $Q+L-1$ nonzero entries since this is just the convolution between the (n,m)th CIR and the UW. Since there are ML unknown CIR coefficients, this results in the necessary (but not sufficient) condition for channel identifiability $Q + L - 1 \geq ML$, which is perhaps better expressed as

$$M \leq \frac{Q-1}{L} + 1. \tag{11}$$

In practical systems, the UW must be at least as long as the memory order of the CIR (i.e., $Q \geq L-1$) in order to induce circularity in the channel and facilitate low-complexity frequency-domain equalization at the receiver. Furthermore, the UW should be designed such that it can support channel

estimation for a given (maximum) delay spread while occupying a minimal amount of overhead.[2] Consequently, it follows that the UW should be chosen to be on the order of the discrete channel length L. By choosing $Q = L - 1$, it is apparent from (11) that only one transmit antenna can be supported while maintaining channel identifiability. However, by increasing the UW overhead to $Q = L + 1$, two transmit antennas can be supported. When $L \gg 2$, this additional overhead is very small. Note that in order to maintain channel identifiability for $M > 2$ transmit antennas, Q must be increased by L samples per additional antenna, which leads to a large overhead.

2.2. MIMO OFDM with nulled subcarriers

In this section, a MIMO OFDM system with a preamble consisting of a number of OFDM symbols used for training is considered, and some subcarriers in this system are nulled. This problem was first considered in [19] where it was shown that conventional MIMO OFDM training schemes are not necessarily optimal when subcarriers are nulled, which is always the case in practice. In [34], a method of constructing optimal preambles for OFDM systems with nulled subcarriers was presented; however, it was also shown that this method is only viable when $S \geq M(2L - 1)$ where S is the number of active subcarriers in the preamble. It will be shown below that the method proposed in this paper relaxes this bound to $S \geq ML$.

A block diagram of a MIMO OFDM system is illustrated in Figure 3. Much of the notation that was used in Section 2.1 to describe a MIMO UW system will be employed here, and it will soon become apparent that MIMO UW and MIMO OFDM systems can be described mathematically by using

---

[2] This approach is in contrast to the method discussed in [31] where the UW is designed specifically for channel estimation, in which case the amount of overhead that is required is not considered an important issue.

very similar approaches. Throughout this discussion, it is assumed that the channel is constant for the duration of a packet, but varies from packet to packet. The CP in each OFDM symbol converts the linear convolution of the channel into cyclic convolution; hence, the input-output relationship of the system can be described in a similar manner to the MIMO UW case, where the post-DFT block of symbols for the S active subcarriers in the system at the nth receive antenna is given by

$$\widetilde{\mathbf{y}}_n = \sum_{m=1}^{M} \mathbf{H}_{n,m} \widetilde{\mathbf{x}}_m + \widetilde{\mathbf{v}}_n, \tag{12}$$

where $\mathbf{H}_{n,m}$ is the $S \times S$ diagonal matrix of the frequency response coefficients for the active subcarriers in the (n,m)th channel, $\widetilde{\mathbf{x}}_m$ is the length-S active data (or training) signal at the mth transmit antenna specified in the frequency domain, and $\widetilde{\mathbf{v}}_n$ is a vector of zero-mean, white Gaussian noise samples with variance $\sigma_v^2/2$ per dimension. This system expression can be rewritten as

$$\widetilde{\mathbf{y}}_n = \sum_{m=1}^{M} \sqrt{K} \widetilde{\mathbf{X}}_m \mathbf{W} \mathbf{g}_{n,m} + \widetilde{\mathbf{v}}_n = \mathbf{B} \mathbf{g}_n + \widetilde{\mathbf{v}}_n, \tag{13}$$

where $\mathbf{W} \in \mathbb{C}^{S \times L}$ is a partial DFT matrix choosing the S active subcarriers and the L time domain channel taps, $\mathbf{B} := \sqrt{K}(\widetilde{\mathbf{X}}_1 \mathbf{W}, \ldots, \widetilde{\mathbf{X}}_M \mathbf{W})$, and $\widetilde{\mathbf{X}}_m := \mathrm{diag}\{\widetilde{\mathbf{x}}_m\}$. It is assumed that $S \geq ML$; thus, the matrix $\mathbf{B}$ is tall and has full column rank. Note that this condition is similar to the condition for channel identifiability stated for the UW case in Section 2.1. It follows that the LS channel estimate is given by

$$\widehat{\mathbf{g}}_n = \left(\mathbf{B}^H \mathbf{B}\right)^{-1} \mathbf{B}^H \widetilde{\mathbf{y}}_n. \tag{14}$$

As in the previous section, it is obvious that the channel estimate is independent of the receive antenna; consequently, the index n can be omitted.

As with MIMO UW systems, OFDM systems exploit the frequency domain to perform channel equalization. Consequently, the frequency domain estimate of the channel is of more interest than the estimate of the CIR. This estimate is given by

$$\widehat{\mathbf{h}} = \left(\mathbf{I}_M \otimes \mathbf{W}\right) \widehat{\mathbf{g}}. \tag{15}$$

Note that this is only a partial channel estimate, where the frequency response coefficients have been estimated for the active subcarriers only. The MSE of this LS channel estimate is given by

$$\mathrm{MSE} = \sigma_v^2 \mathrm{Tr}\left\{\left(\mathbf{I}_M \otimes \mathbf{W}\right)\left(\mathbf{B}^H \mathbf{B}\right)^{-1}\left(\mathbf{I}_M \otimes \mathbf{W}\right)^H\right\} \tag{16}$$

which is minimized when the matrix $\mathbf{B}$ is unitary [10, 19].

When no subcarriers are nulled (i.e., $S = K$), sequences can be easily designed such that this condition is met [10, 12, 19]. However, it was shown in [19] that nulling subcarriers causes these conventional optimal sequences to be suboptimal in many cases. As with MIMO UW design, nonlinear optimization techniques can be employed to find sequences that come close to minimizing the MSE of the channel estimate when nulled subcarriers are used.

3. NUMERICAL OPTIMIZATION WITH CONSTRAINTS

Due to their computationally complex nature, the optimization problems stated above cannot be solved analytically, or are at least intractable. However, numerical methods can be applied to solve these problems with good results. In this section, standard nonlinear optimization techniques are briefly reviewed. In particular, one such technique known as the *barrier method* is discussed and its application to the MIMO training sequence optimization problem is detailed. Furthermore, practical constraints such as the mean power and the peak power of the sequences are discussed in the context of the optimization problem; these constraints can be easily added to the problem when the barrier method is employed.

3.1. Standard optimization techniques

Constrained optimization problems generally are of the form [22, 35]

$$\begin{aligned} \text{minimize } & f_0(\mathbf{z}), \\ \text{subject to } & f_i(\mathbf{z}) \leq 0, \quad i = 1, \ldots, p, \\ & r_i(\mathbf{z}) = 0, \quad i = 1, \ldots, q, \end{aligned} \tag{17}$$

where $\mathbf{z}$ is the optimization variable (in this case, the UWs or the OFDM training sequences), f_0 is the objective or cost function, f_i are inequality constraints, and r_i are equality constraints. Note that f_0, f_i, and r_i are all real-valued scalar functions of a complex vector $\mathbf{z}$. If the objective function and the inequality constraint functions are convex, and the equality constraint functions are linear, then the theory of convex optimization can be used to solve this problem.

Convex optimization is a well-researched field; its popularity owing largely to the fact that most convex problems can be solved efficiently [22]. When convex optimization problems cannot be solved analytically, which is often the case, one must resort to various numerical methods, such as steepest descent algorithms or Newton's method. The latter of these two techniques is generally very efficient at solving problems with equality constraints only. However, when inequality constraints are introduced, other techniques such as IP methods must be employed. IP methods solve the convex optimization problem given by (17) by employing Newton's method to solve a sequence of equality constrained (or unconstrained) subproblems. Even when a problem is not convex, IP methods can sometimes be used to great effect (see, e.g., [25] and the references therein).

One popular IP method that can be used to solve nonlinear optimization problems is the barrier method. This method is documented for convenience in Algorithm 1 [22]. By applying the barrier method, the problem given in (17) can be restated as

$$\begin{aligned} \text{minimize } & f_0(\mathbf{z}) + \sum_{i=1}^{p} I\left(f_i(\mathbf{z})\right), \\ \text{subject to } & r_i(\mathbf{z}) = 0, \quad i = 1, \ldots, q, \end{aligned} \tag{18}$$

where $I : \mathbb{R} \to \mathbb{R}$ is the indicator function for nonpositive

given strictly feasible $\mathbf{z}$, $t > 0$, $\mu > 1$, $\epsilon_o > 0$,
$\epsilon_i > 0$
repeat
   (1) Newton's method ($\mathbf{z}$, $\epsilon_i > 0$)
      (a) $\Delta\mathbf{z} = -\nabla^2 f(\mathbf{z})^{-1}\nabla f(\mathbf{z})$
          $\lambda^2 = -\nabla f(\mathbf{z})^H \Delta\mathbf{z}$
      (b) quit if $\lambda^2/2 < \epsilon_i$
          return $\mathbf{z}^* := \mathbf{z}$
      (c) Line search (determine β)
      (d) $\mathbf{z} := \mathbf{z} + \beta\Delta\mathbf{z}$
   (2) $\mathbf{z} := \mathbf{z}^*$
   (3) quit if $p/t < \epsilon_o$
   (4) $t := \mu t$

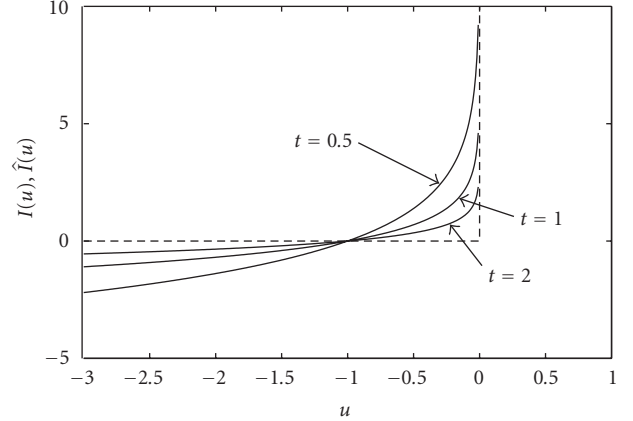ALGORITHM 1: The barrier method.

FIGURE 4: Indicator function and approximate logarithmic functions. The dashed lines show the indicator function and the solid lines show the approximations for $t = 0.5, 1, 2$. The best approximation is given by $t = 2$ [22].

TABLE 1: Differences in MSE expressions for MIMO UW and MIMO OFDM channel estimates.

MIMO UW		MIMO OFDM
$\mathbf{F}_L$	$\Longleftrightarrow$	$\mathbf{W}$
$\widetilde{\mathbf{U}}_m := \mathrm{diag}\{\mathbf{F}'_Q\mathbf{u}_m\}$	$\Longleftrightarrow$	$\widetilde{\mathbf{X}}_m := \mathrm{diag}\{\bar{\mathbf{x}}_m\}$

real numbers given by

$$I(u) = \begin{cases} 0, & u \leqslant 0, \\ \infty, & u > 0. \end{cases} \tag{19}$$

The indicator function can, in practice, be approximated by the function

$$\widehat{I}(u) = -\frac{1}{t}\log(-u), \tag{20}$$

where t is the logarithmic barrier accuracy parameter and (by convention) $\widehat{I}(u) = \infty$ for $u > 0$. Figure 4 illustrates the indicator function and its approximation for several values of t. When the equality constraints r_i shown in (18) are linear or do not exist, Newton's method can be used to find an optimal point $\mathbf{z}^*$ over the search space as outlined in Algorithm 1. Note that typical values of the tolerances ϵ_o and ϵ_i, which are shown in Algorithm 1, are in the region $0.001 \leq \epsilon_o, \epsilon_i \leq 0.1$, and the scaling factor μ is generally chosen such that $10 \leq \mu \leq 20$. Also, it is worth noting that the parameters t and p used in Algorithm 1 are the logarithmic barrier accuracy parameter and the number of constraints, respectively.

An alternative to the barrier method is the *primal-dual* IP method. The primal-dual method is similar to the barrier method in a number of ways. In general, the only differences between the two techniques lie with the search directions, the loop structure of the algorithm (the primal-dual method only has one loop), and the fact that the temporary solutions with each iteration of the primal-dual method are not necessarily feasible (i.e., they may not meet the constraints of the problem) [22]. For brevity, only the barrier method will be used in this paper.

3.2. Reformulating the MSE for the barrier method

The barrier method requires the objective function to be twice differentiable with respect to the optimization variable. In the examples discussed in this paper, the objective func-

tion is the MSE of the channel estimate and the optimization variable is the set of training sequences or UWs. Consequently, it is beneficial to reformulate the expressions for MSE given by (10) and (16) to be functions of a single vector of UWs or training sequences. Expressing the problem in this form facilitates simple differentiation of the objective functions through the derivation of gradients and Hessians of the functions.

Notice that the MSE expressions given by (10) and (16) are very similar. In fact, the structures of the two expressions are identical. The only differences lie with the definitions of the partial DFT matrix and the training signal.[3] These differences are outlined in Table 1. Due to the similarities of the two MSE expressions, a single general expression for the MSE that encompasses the two examples discussed in Sections 2.1 and 2.2 can be derived. This general formula for the MSE is given by

$$\mathrm{MSE}(\mathbf{z}) \propto \mathrm{Tr}\big\{\big(\mathbf{I}_M \otimes \mathbf{\Psi}\mathbf{F}_L\big)\big(\big(\mathbf{I}_{ML} \otimes \mathbf{\Phi}\mathbf{z}\big)^H \\ \times \mathbf{J}\big(\mathbf{I}_{ML} \otimes \mathbf{\Phi}\mathbf{z}\big)\big)^{-1}\big(\mathbf{I}_M \otimes \mathbf{\Psi}\mathbf{F}_L\big)^H\big\}, \tag{21}$$

where $\mathbf{z}$ is a stacked column vector of training sequences or UWs, $\mathbf{J}$ is a sparse matrix that contains elements of the DFT matrix, and $\mathbf{\Psi}$ and $\mathbf{\Phi}$ are defined differently according to whether UW optimization is being performed for single-carrier MIMO systems or training sequence optimization is

---

[3] The noise variance scaling parameters in (10) and (16) are ignored here since they have no bearing on the optimal design of the training sequences.

being performed for MIMO OFDM systems with nulled sub-carriers. Consider the example where MIMO UW optimization is performed. In this case, $\mathbf{z} := (\mathbf{u}_1^T, \ldots, \mathbf{u}_M^T)^T$, $\boldsymbol{\Psi} := \mathbf{I}_K$, $\boldsymbol{\Phi} := \mathbf{I}_M \otimes \mathbf{F}_Q'$, and $\mathbf{J} \in \mathbb{C}^{M^2KL \times M^2KL}$ contains elements of $\mathbf{F}_L$. For the example where MIMO OFDM training sequence optimization is performed, $\mathbf{z} := (\tilde{\mathbf{x}}_1^T, \ldots, \tilde{\mathbf{x}}_M^T)^T$, $\boldsymbol{\Psi} \in \{0, 1\}^{S \times K}$ is defined as the S rows of $\mathbf{I}_K$ corresponding to the S active subcarriers, $\boldsymbol{\Phi} := \mathbf{I}_{MS}$, and $\mathbf{J} \in \mathbb{C}^{M^2SL \times M^2SL}$ contains elements of $\mathbf{W}$. The full details of the reformulation of the expression for the MSE of a MIMO channel estimate can be found in Appendix A.

Note that the proportionality in (21) does not affect the minimization of the MSE. Consequently, the expression given on the right-hand side of (21) can be directly used as the objective function $f_0(\mathbf{z})$ in the minimization problem stated in (17). Furthermore, this function is twice differentiable, which is a requirement of the barrier method. The gradient and the Hessian of this function are given in Appendix B.

3.3. Constraints on MIMO training sequences

In order to obtain meaningful results from the optimization algorithm, a mean power constraint must be placed on the training sequences and UWs. Without this constraint, the optimization algorithm would simply increase the power of the sequences with each iteration, which would obviously lead to a lower channel estimation MSE. It is desirable to make the mean power constraint an equality constraint, such as $\|\mathbf{z}\|^2 = 1$. Unfortunately, Newton's method, and thus the barrier method, do not support quadratic equality constraints [22]. A small tolerance ε (say, $\varepsilon = 0.01$) can be added to an inequality constraint to circumvent this problem, giving the constraint

$$1 - \varepsilon \leq \|\mathbf{z}\|^2 \leq 1 + \varepsilon. \tag{22}$$

Note that all solutions to the optimization problem can be normalized to have the same power without significantly affecting the optimality of the sequences. By defining the logarithmic constraint function as[4] $\phi_i(\mathbf{z}) = -\log(-f_i(\mathbf{z}))$, the logarithmic mean power constraints can be expressed as

$$\begin{aligned} \phi_1(\mathbf{z}) &= -\log\left(1 + \varepsilon - \|\mathbf{z}\|^2\right), \\ \phi_2(\mathbf{z}) &= -\log\left(\|\mathbf{z}\|^2 - 1 + \varepsilon\right). \end{aligned} \tag{23}$$

Another desirable property of wireless transmissions, whether for training or data transfer, is that they have a low PAPR. The PAPR of the training sequences (or UWs) can be limited by employing a peak power constraint in addition to the mean power constraint discussed above. The constraint on the peak power of the transmitted signal can be written as

$$\left| \mathbf{e}_i^T \boldsymbol{\Theta} \mathbf{z} \right|^2 \leq \delta, \quad \forall i, \tag{24}$$

---

[4] The multiplication of the objective and constraint functions by t does not alter the optimization problem.

where the matrix $\boldsymbol{\Theta}$ defines the mapping of the data vector $\mathbf{z}$ to the time domain and $\mathbf{e}_i$ is the ith unit vector of the appropriate size. In practical systems, the PAPR constraint should be applied to the oversampled signal [36]. Consequently, $\boldsymbol{\Theta}$ must account for filtering or interpolation between time-domain samples. Many different filtering strategies exist, but a common approach is to use a raised cosine filter [37]. Using this approach, the mapping matrix can be defined as

$$\boldsymbol{\Theta}_{\mathrm{UW}} := \mathbf{I}_M \otimes \mathbf{C} \tag{25}$$

for the UW case, where $\mathbf{C}$ is the $\rho Q \times Q$ raised cosine filter matrix. In the case of the OFDM sequences, the mapping matrix should be defined as

$$\boldsymbol{\Theta}_{\mathrm{OFDM}} := \mathbf{I}_M \otimes \mathbf{C}\overline{\mathbf{W}}^H, \tag{26}$$

where $\overline{\mathbf{W}} \in \mathbb{C}^{S \times K}$ is the normalized DFT matrix mapped to the S active subcarriers and $\mathbf{C}$ is the $\rho K \times K$ raised cosine filter matrix. Although the size of $\mathbf{C}$ varies for the two cases, the (i, k)th element of $\mathbf{C}$ is defined as

$$C_{i,k} = \mathrm{sinc}\left(\pi \frac{(i - \rho k)}{\rho}\right) \frac{\cos\left(\pi\alpha((i - \rho k)/\rho)\right)}{1 - \left(2\alpha((i - \rho k)/\rho)\right)^2} \tag{27}$$

for both cases, where $0 \leq \alpha \leq 1$ is the roll-off factor. Regardless of the choice of the mapping matrix $\boldsymbol{\Theta}$, the logarithmic barrier function for the peak power constraint is given by

$$\phi_3(\mathbf{z}) = -\sum_i \log\left(\delta - \left|\mathbf{e}_i^T \boldsymbol{\Theta} \mathbf{z}\right|^2\right). \tag{28}$$

Notice that the three logarithmic constraint functions given above are twice differentiable. The gradients and Hessians of these functions can be found in Appendix C. By using these constraint functions, the optimization problem can be rewritten as

$$\text{minimize } f(\mathbf{z}) = t f_0(\mathbf{z}) + \sum_{i=1}^{3} \phi_i(\mathbf{z}) \tag{29}$$

which can be solved by employing the barrier method as described in Algorithm 1.

3.4. Issues of convergence

As previously mentioned, the barrier method works well when the objective and constraint functions are convex. Unfortunately, this is not the case with the two examples discussed in this paper; indeed, the objective function given by (21) is not convex, which can be shown through a numerical counterexample. Consequently, there exist local minima that are not equal to the global minimum. The barrier method can be employed to find *a* solution to this optimization problem, but it may not be the *optimal* solution. The purpose of using this technique, however, is to find *near*-optimal sequences, which may or may not be the best possible sequences that exist under the given constraints. Consequently, it is usually enough to find a sequence that converges to a low

local minimum since, as it will be shown later through experimental results, these minima are generally low enough to provide near-optimal performance in the MSE sense.

One way of ensuring that a good sequence is found is to use several different (possibly random) feasible starting vectors.[5] If a large number of feasible starting vectors is used, the likelihood that the barrier method will converge to a low local minimum, or indeed the global minimum, is high. A similar technique was used in [25]. It should be noted that the complexity of computing multiple "optimal" sequences is not a significant issue since this can be done offline and the best results can be stored for future use.

4. SIMULATION RESULTS

In this section, results obtained through computer simulations are shown. These results depict the benefits that can be gained by employing nonlinear optimization techniques to design MIMO training sequences under difficult constraints. Furthermore, characteristics of the near-optimal sequences are discussed. In particular, the structure of the sequences generated by the proposed approach and the trade-off between the PAPR of the sequences and the achievable MSE of the channel estimate are investigated. Results are given for both the MIMO UW scenario and the MIMO OFDM scenario.

4.1. Channel model and assumptions

The training strategies discussed in this paper are particularly suitable for use in wireless local area networks (WLANs) where the Doppler spread is low (on the order of a few Hz). Consequently, the IEEE 802.11n channel models [38] are used to obtain the results presented below. These models are cluster-based and cover six fundamental cases ranging from model "A" (frequency-flat) to model "F" (150 nanoseconds root-mean square (RMS) delay spread). In the following discussion, the bandwidth of each transmission is 20 MHz at a center frequency of 5.2 GHz, and each block (for both single-carrier and multicarrier systems) comprises $K = 64$ symbols and has a guard interval of 16 samples. Thus, the coherence time of the channel is several orders of magnitude greater than the period of a transmitted block ($4\,\mu s$). As a result, quasistatic fading is assumed in the following scenarios.

4.2. MIMO UW

One interesting, and intuitively satisfying, result of MIMO training sequence design is that the PAPR of the training sequences cannot be decreased without compromising the MSE of the channel estimate. This trade-off can be observed for the MIMO UW case in Figure 5, where the system in question has $M = 2$ transmit antennas, a UW length of $Q = 16$ symbols, and a block size of $K = 64$. A raised cosine filter with a roll-off factor of $\alpha = 0.2$ and an oversam-
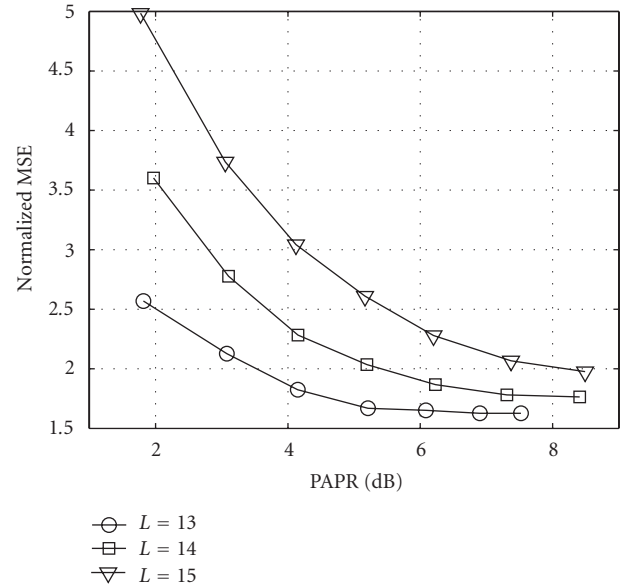

FIGURE 5: Normalized MSE versus PAPR for three different lengths of CIR in a UW system.

pling factor of $\rho = 4$ was used. As shown in this example, the normalized MSE of the channel estimate, which is defined as

$$\overline{\text{MSE}} = \frac{1}{\widetilde{\sigma}^2 MK}\text{MSE}, \tag{30}$$

where MSE is given by (10), is smaller for shorter CIRs. This behavior is due to the time-domain windowing performed by the LS channel estimator, which reduces the noise in the channel estimate. It is worth noting that all curves level out to a point beyond which increasing the allowed PAPR does not reduce the MSE any further. To put these results into perspective, the 99.99 percentile PAPRs for QPSK, 16-QAM, and 64-QAM signals are 5.7 dB, 6.8 dB, and 7.1 dB, respectively.

To investigate the impact that block averaging has on bit-error rate (BER), a UW system with $M = 2$ transmit antennas and $N = 2$ receive antennas, a UW of length $Q = 16$ (designed for a channel of length $L = 15$), a block size of $K = 64$ symbols, and a CIR based on the IEEE 802.11n channel model B [38], which is a model of an indoor environment with 15 nanoseconds RMS delay spread and 10 Hz RMS Doppler spectrum spread, was simulated. QPSK signaling was employed, and the packet size was varied from three block intervals to 50 block intervals (i.e., six to 100 blocks in total). A rate-1/2, memory-6 convolutional code was used, and the receiver employed a linear MMSE frequency-domain equalizer. Note that the 99.99 percentile PAPR of a QPSK signal in this example is 5.7 dB. Consequently, the UWs used in this example were constrained to have a PAPR less than 5.7 dB. The results of this simulation are plotted in Figure 6. The system using optimized UWs (labeled "optimized UW") and block averaging as described in Section 2.1 was compared to a system using a one-block preamble supporting both antennas (labeled "1 preamble") [12] as well as to a system using time-multiplexed

---

[5] A *feasible* vector is defined as a vector that satisfies the inequality constraints in the optimization problem.
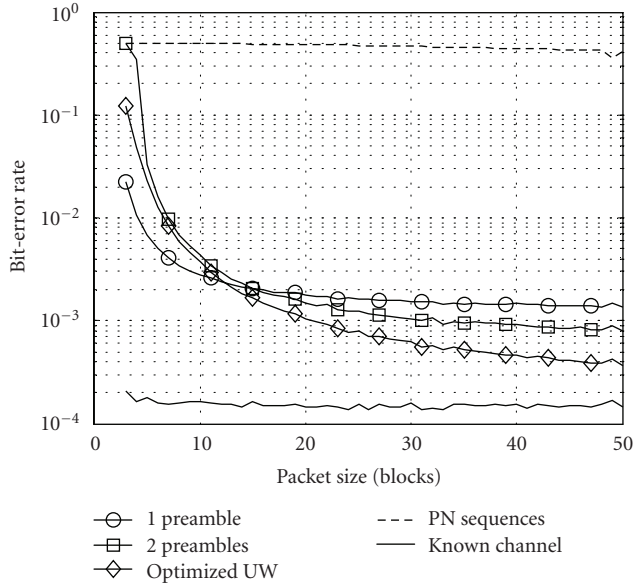
FIGURE 6: BER versus packet length for five different systems: "1 preamble" uses a single preamble, which supports both transmit antennas; "2 preambles" uses time-multiplexed preambles; "optimized UW" uses optimized UWs to estimate the channel; "PN sequences" uses PN sequences to estimate the channel; and "known channel" has perfect knowledge of the channel state information. $M = N = 2$, $K = 64$, $Q = 16$, $L = 15$, SNR = 20 dB with a rate-1/2 convolutional code.

preambles (labeled "2 preambles"). The two latter systems employ puncturing to achieve the same packet size as the former system (see, e.g., Figure 7). Thus, as the packet length increases, the puncturing is less severe for these two systems. Also, a system that uses pseudonoise (PN) sequences as UWs was simulated, and a system with perfect knowledge of the channel was simulated as a reference. As shown in Figure 6, the system using optimized UWs and block averaging to perform channel estimation performs poorly for short packets. However, for packets consisting of fifteen block intervals (approximately 1500 bits) or more, the block averaging system outperforms the two systems that use preambles. The system that utilizes block averaging and PN sequences performs poorly for all simulated packet lengths.

4.3. MIMO OFDM with nulled subcarriers

In this section, an OFDM system based on the IEEE 802.11a specification [39] is considered. IEEE 802.11a systems employ $K = 64$ subcarriers with $S = 52$ carrying data where the nulled subcarriers are defined by the set $\{0, 27, \ldots, 37\}$. In [19], it was noted that for two transmit antennas, simple sequences such as $\tilde{\mathbf{x}}_1 = (1, 0, 1, \ldots, 1, 0)^T$ and $\tilde{\mathbf{x}}_2 = (0, 1, 0, \ldots, 0, 1)^T$ (i.e., transmitting on alternate subcarriers) perform well, although they do not meet the lower bound. However, this alternate subcarrier transmission strategy does not generally perform well when $M > 2$. In this section, similar sequences are used as a reference point to show that it is

possible to design better sequences using the barrier method. In Figure 8, the normalized MSE of the channel estimate, which is defined as

$$\overline{\text{MSE}} = \frac{1}{\sigma_v^2 MK}\text{MSE}, \qquad (31)$$

where MSE is given by (16), is plotted as a function of the channel length L for the reference cases described above and for the case where the sequences are designed as discussed in Section 3. The number of transmit antennas is set to $M = 3$ in this example. As observed in Figure 8, the designed sequences have a distinct advantage over the alternating subcarriers at large channel lengths; whereas, both sets of sequences are very close to the lower bound for shorter channels.

The BER of an OFDM system that utilizes $M = N = 3$ transmit and receive antennas and optimized preambles is depicted in Figure 9. In this example, the BER is shown for IEEE 802.11n channel model E [38], which has an excess delay spread of 750 nanoseconds ($L = 15$ samples). The system uses 16-QAM modulation and a rate-3/4, memory-6 convolutional code. It is observed that when one or two OFDM symbols are used for the preamble, the system employing the sequences that were designed through nonlinear optimization techniques performs 2-3 dB better than the system that transmits training on alternating subcarriers, which is the optimal case when no nulled subcarriers are employed. It should also be noted that the method proposed in [34] cannot be used to find optimal sequences in this example since $S < M(2L - 1)$.

4.4. Sequence structure

It is interesting to observe the structure of the sequences that are generated by the proposed optimization algorithm. It was found that the phases of the sequence elements appear to be random, both for the OFDM training sequences and the single-carrier UWs. Similarly, the envelopes of the OFDM training sequences (in the frequency domain) generally have no clear structure apart from the location of the nulled subcarriers, which are common to all sequences. However, the near-optimal UWs are more structured. In particular, the envelopes of all of the UWs that were designed by the proposed method exhibit a distinctive trough in the center, with peaks occurring near the edges. The depth of this trough (and thus the heights of the peaks) obviously depends upon the PAPR constraint that was employed to generate the sequences, but for a constraint of greater than 4 dB, the deepest point on the trough is typically close to zero and most of the energy in the UWs is contained in the first and last few elements. As an illustration of this phenomenon, the powers of an OFDM training sequence waveform and a single-carrier UW—in particular, two of the sequences that were employed to produce the results shown in Sections 4.2 and 4.3—are depicted in Figure 10. The trough can clearly be seen in the plot of the UW waveform, whereas the time-domain OFDM waveform appears to have no recognizable structure. It should be noted that the properties exhibited by
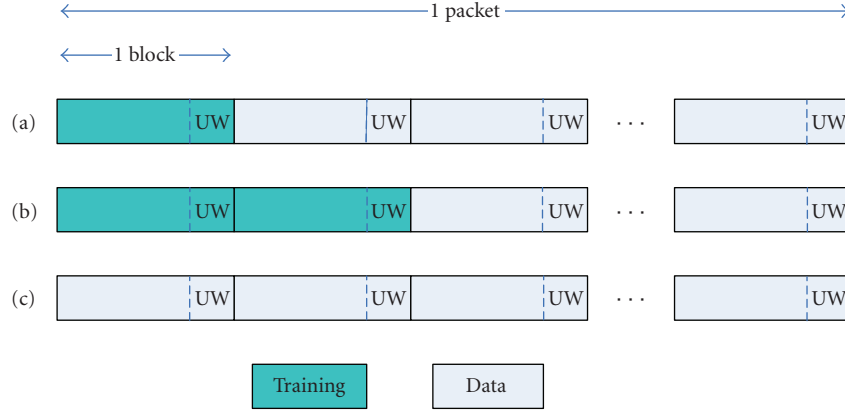
FIGURE 7: Illustration of packet format for MIMO UW simulations with puncturing: (a) 1 preamble, (b) 2 preambles, and (c) UW only.
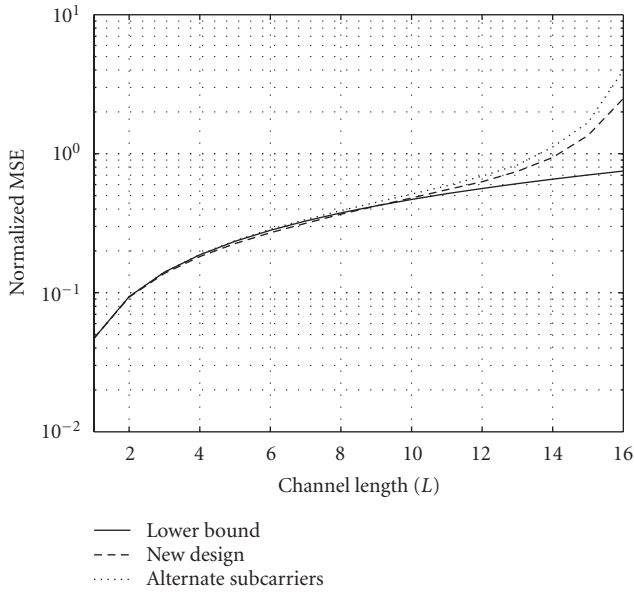

FIGURE 8: Normalized MSE for OFDM systems with $M = 3$.

FIGURE 9: BER versus SNR for a 3×3 OFDM system: 16-QAM, rate-3/4 convolutional code, IEEE 802.11n channel model E.

the waveforms shown in Figure 10 are characteristics of all OFDM training sequences and single-carrier UWs generated by the proposed optimization method.

5. CONCLUSIONS

In this paper, nonlinear optimization techniques were used to design near-optimal sequences for MIMO channel estimation. In particular, two example scenarios were explored: single-carrier MIMO transmissions with a UW extension and MIMO OFDM transmissions with nulled subcarriers. A generalized expression for the MSE of the LS channel estimate was given as a function of a single vector of training sequences. This expression was used along with the barrier IP method to find near-optimal sequences with a constrained PAPR. The advantages of using the optimized sequences were d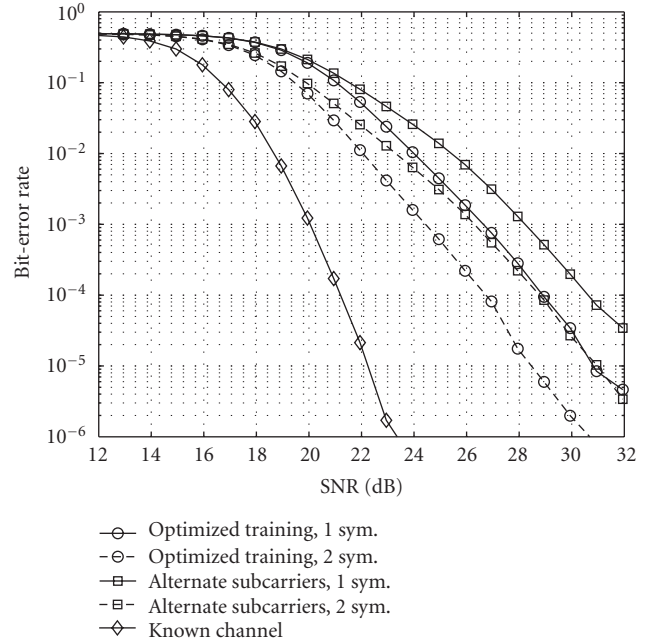emonstrated by computing both the MSE and the BER of various systems through computer simulations. The new sequences were shown to provide better channel estimates than conventional sequences for all of the systems that were investigated. It should be noted that the techniques presented in this paper can be performed offline since training sequences are generally specified by the system designer.

APPENDICES

A. REFORMULATION OF MSE COST FUNCTION

A generalized equation for the MSE of a MIMO channel estimate for the two examples discussed in this paper can be derived by adopting the matrix definitions given in Section 3.2
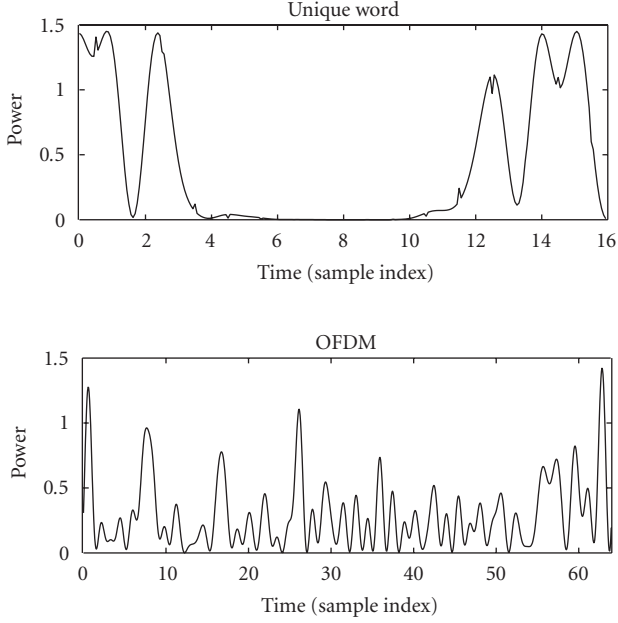
FIGURE 10: Time-domain waveforms for an optimized UW and an optimized OFDM training sequence.

and observing the relations given in Table 1. This expression can be written as a function of a single concatenated training sequence, or UW, vector $\mathbf{z}$ in order to facilitate the application of the nonlinear optimization techniques discussed in Section 3. This reformulation of the MSE as a function of a vector is best described through an example; the generalization follows easily.

Consider the example of the OFDM system. The MSE in (16) depends on $\mathbf{B}^H\mathbf{B}$, which is an $M \times M$ block matrix where each submatrix is of size $L \times L$ and is given by $\mathbf{W}^H\widetilde{\mathbf{X}}_m^H\widetilde{\mathbf{X}}_{m'}\mathbf{W}$. After performing some algebraic manipulations, this expression for the (m, m')th sub-matrix can be written as

$$
\mathbf{W}^H\widetilde{\mathbf{X}}_m^H\widetilde{\mathbf{X}}_{m'}\mathbf{W}
$$
$$
= (\mathbf{I}_L \otimes \widetilde{\mathbf{x}}_m)^H \begin{pmatrix} \boldsymbol{\Omega}_{0,0} & \cdots & \boldsymbol{\Omega}_{0,L-1} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Omega}_{L-1,0} & \cdots & \boldsymbol{\Omega}_{L-1,L-1} \end{pmatrix} (\mathbf{I}_L \otimes \widetilde{\mathbf{x}}_{m'}),
$$
(A.1)

where $\boldsymbol{\Omega}_{\ell,\ell'} = \text{diag}\{W_{0,\ell}W_{0,\ell'}^*, \ldots, W_{S-1,\ell}W_{S-1,\ell'}^*\}$ and $W_{s,\ell}$ is the (s, ℓ)th element of $\mathbf{W}$. The training vectors can be concatenated to form a single vector $\mathbf{z} := (\widetilde{\mathbf{x}}_1^T, \ldots, \widetilde{\mathbf{x}}_M^T)^T$, which can be used to form

$$
\mathbf{B}^H\mathbf{B} = (\mathbf{I}_{ML} \otimes \mathbf{z})^H \underbrace{\begin{pmatrix} \mathbf{J}_{0,0} & \cdots & \mathbf{J}_{0,M-1} \\ \vdots & \ddots & \vdots \\ \mathbf{J}_{M-1,0} & \cdots & \mathbf{J}_{M-1,M-1} \end{pmatrix}}_{\mathbf{J} \in \mathbb{C}^{M^2SL \times M^2SL}} (\mathbf{I}_{ML} \otimes \mathbf{z}),
$$
(A.2)

where $\mathbf{J}_{m,m'} \in \mathbb{C}^{MSL \times MSL}$ is given by

$$
\mathbf{J}_{m,m'} = \begin{pmatrix} (\mathbf{e}_m\mathbf{e}_{m'}^T) \otimes \boldsymbol{\Omega}_{0,0} & \cdots & (\mathbf{e}_m\mathbf{e}_{m'}^T) \otimes \boldsymbol{\Omega}_{0,L-1} \\ \vdots & \ddots & \vdots \\ (\mathbf{e}_m\mathbf{e}_{m'}^T) \otimes \boldsymbol{\Omega}_{L-1,0} & \cdots & (\mathbf{e}_m\mathbf{e}_{m'}^T) \otimes \boldsymbol{\Omega}_{L-1,L-1} \end{pmatrix}
$$
(A.3)

and $\mathbf{e}_m$ is the mth length-M unit vector. The MSE of the OFDM channel estimate can now be written as a function of the vector $\mathbf{z}$ as given by (21) where $\boldsymbol{\Psi}$ is defined such that $\mathbf{W} = \boldsymbol{\Psi}\mathbf{F}_L$ and $\boldsymbol{\Phi}$ is the identity matrix of the appropriate size.

It is easy to see that the generalization of the example given above to include the MIMO UW case is straightforward. In this case, the partial DFT matrix is defined as $\mathbf{F}_L$ rather than $\mathbf{W}$ in the derivation given above. Furthermore, $\mathbf{z} := (\mathbf{u}_1^T, \ldots, \mathbf{u}_M^T)^T$, and the resulting expression for the MSE as a function of $\mathbf{z}$ is given by (21) where $\boldsymbol{\Psi} := \mathbf{I}_K$ and $\boldsymbol{\Phi} := \mathbf{I}_M \otimes \mathbf{F}_Q'$.

B. GRADIENT AND HESSIAN OF THE OBJECTIVE FUNCTION

To derive the gradient and the Hessian of the objective function given by (21), the chain rule and the product rule are required. The chain rule for matrix differentiation is given by $\partial f(\mathbf{X})/\partial t = \text{Tr}\{(\partial f(\mathbf{X})/\partial \mathbf{X})^T(\partial \mathbf{X}/\partial t)\} + \text{Tr}\{(\partial f(\mathbf{X})/\partial \mathbf{X}^*)^T(\partial \mathbf{X}^*/\partial t)\}$ and the product rule is given by $\partial(\mathbf{U}\mathbf{V})/\partial t = (\partial(\mathbf{U})/\partial t)\mathbf{V} + \mathbf{U}(\partial(\mathbf{V})/\partial t)$. By rewriting the MSE in (21) as $\text{MSE}(\mathbf{z}) \propto \text{Tr}\{\mathbf{C}^{-1}\boldsymbol{\Gamma}\}$ where $\boldsymbol{\Gamma} = (\mathbf{I}_M \otimes \boldsymbol{\Psi}\mathbf{F}_L)^H(\mathbf{I}_M \otimes \boldsymbol{\Psi}\mathbf{F}_L)$ and $\mathbf{C} = (\mathbf{I}_{ML} \otimes \boldsymbol{\Phi}\mathbf{z})^H\mathbf{J}(\mathbf{I}_{ML} \otimes \boldsymbol{\Phi}\mathbf{z})$, the ith element of the gradient can be expressed as

$$
\frac{\partial \text{Tr}\{\mathbf{C}^{-1}\boldsymbol{\Gamma}\}}{\partial z_i^*} = \text{Tr}\left\{ -\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial z_i^*}\mathbf{C}^{-1}\boldsymbol{\Gamma} \right\}
$$
(B.1)

which follows from the fact that $\partial \text{Tr}\{\mathbf{U}\}/\partial \mathbf{U} = \mathbf{I}$ and $\partial \mathbf{U}^{-1}/\partial t = -\mathbf{U}^{-1}(\partial \mathbf{U}/\partial t)\mathbf{U}^{-1}$. The inner derivative is $\partial \mathbf{C}/\partial z_i^* = (\mathbf{I}_{ML} \otimes \mathbf{e}_i)^T\mathbf{J}(\mathbf{I}_{ML} \otimes \boldsymbol{\Phi}\mathbf{z})$, where $\mathbf{e}_i$ is the ith unit vector of the appropriate size; consequently, the ith element of the gradient is given by

$$
\frac{\partial \text{Tr}\{\mathbf{C}^{-1}\boldsymbol{\Gamma}\}}{\partial z_i^*} = \text{Tr}\{ -\mathbf{C}^{-1}\boldsymbol{\Gamma}\mathbf{C}^{-1}(\mathbf{I}_{ML} \otimes \mathbf{e}_i)^T\mathbf{J}(\mathbf{I}_{ML} \otimes \boldsymbol{\Phi}\mathbf{z}) \}.
$$
(B.2)

The Hessian can be defined as the matrix of derivatives of the conjugate of (B.2). Note that $(\text{Tr}\{\mathbf{A}\})^* = \text{Tr}\{\mathbf{A}^H\}$ and $\mathbf{C}, \boldsymbol{\Gamma}$, and $\mathbf{J}$ are Hermitian symmetric. Now, by defining $\boldsymbol{\Xi} = -\mathbf{C}^{-1}\boldsymbol{\Gamma}\mathbf{C}^{-1}$ and $\boldsymbol{\Lambda} = (\mathbf{I}_{ML} \otimes \boldsymbol{\Phi}\mathbf{z})^H\mathbf{J}(\mathbf{I}_{ML} \otimes \mathbf{e}_i)$ and using the chain and product rules, the (i', i)th element of the Hessian matrix can be written as

$$
\frac{\partial}{\partial z_{i'}^*}\left(\frac{\partial \text{Tr}\{\mathbf{C}^{-1}\boldsymbol{\Gamma}\}}{\partial z_i^*}\right)^* = \text{Tr}\left\{ \frac{\partial(\boldsymbol{\Xi}\boldsymbol{\Lambda})}{\partial z_{i'}^*} \right\} = \text{Tr}\left\{ \frac{\partial \boldsymbol{\Xi}}{\partial z_{i'}^*}\boldsymbol{\Lambda} + \boldsymbol{\Xi}\frac{\partial \boldsymbol{\Lambda}}{\partial z_{i'}^*} \right\},
$$
(B.3)

where $\partial \boldsymbol{\Xi}/\partial z_{i'}^* = -(\partial \mathbf{C}^{-1}/\partial z_{i'}^*)\boldsymbol{\Gamma}\mathbf{C}^{-1} - \mathbf{C}^{-1}\boldsymbol{\Gamma}(\partial \mathbf{C}^{-1}/\partial z_{i'}^*)$ and $\partial \boldsymbol{\Lambda}/\partial z_{i'}^* = (\mathbf{I}_{ML} \otimes \mathbf{e}_{i'})^H\mathbf{J}(\mathbf{I}_{ML} \otimes \mathbf{e}_i)$. Using the inner derivative

$\partial \mathbf{C}^{-1}/\partial z_{i'}^* = -\mathbf{C}^{-1}(\mathbf{I}_{ML} \otimes \mathbf{e}_{i'})^H \mathbf{J}(\mathbf{I}_{ML} \otimes \mathbf{\Phi z})\mathbf{C}^{-1}$, the Hessian can finally be written as

$$
\frac{\partial}{\partial z_{i'}^*}\left(\frac{\partial \mathrm{Tr}\{\mathbf{C}^{-1}\mathbf{\Gamma}\}}{\partial z_i^*}\right)^*
$$
$$
= \mathrm{Tr}\big\{(\mathbf{C}^{-1}(\mathbf{I}_{ML} \otimes \mathbf{e}_{i'})^H \mathbf{J}(\mathbf{I}_{ML} \otimes \mathbf{\Phi z})\mathbf{C}^{-1}\mathbf{\Gamma}\mathbf{C}^{-1}
$$
$$
\quad + \mathbf{C}^{-1}\mathbf{\Gamma}\mathbf{C}^{-1}(\mathbf{I}_{ML} \otimes \mathbf{e}_{i'})^H \mathbf{J}(\mathbf{I}_{ML} \otimes \mathbf{\Phi z})\mathbf{C}^{-1}) \quad \text{(B.4)}
$$
$$
\quad \times (\mathbf{I}_{ML} \otimes \mathbf{\Phi z})^H \mathbf{J}(\mathbf{I}_{ML} \otimes \mathbf{e}_i)
$$
$$
\quad - \mathbf{C}^{-1}\mathbf{\Gamma}\mathbf{C}^{-1}(\mathbf{I}_{ML} \otimes \mathbf{e}_{i'})^H \mathbf{J}(\mathbf{I}_{ML} \otimes \mathbf{e}_i)\big\}.
$$

C. GRADIENTS AND HESSIANS OF THE CONSTRAINT FUNCTIONS

The gradients and Hessians of the constraints are straightforward to find and are given by

$$
\nabla \phi_1(\mathbf{z}) = \frac{2}{1+\varepsilon - \|\mathbf{z}\|^2}\mathbf{z},
$$
$$
\nabla^2 \phi_1(\mathbf{z}) = \frac{2}{(1+\varepsilon - \|\mathbf{z}\|^2)^2}(2\mathbf{z}\mathbf{z}^H + (1+\varepsilon - \|\mathbf{z}\|^2)\mathbf{I}),
$$
$$
\nabla \phi_2(\mathbf{z}) = \frac{2}{1-\varepsilon - \|\mathbf{z}\|^2}\mathbf{z},
$$
$$
\nabla^2 \phi_2(\mathbf{z}) = \frac{2}{(1-\varepsilon - \|\mathbf{z}\|^2)^2}(2\mathbf{z}\mathbf{z}^H + (1-\varepsilon - \|\mathbf{z}\|^2)\mathbf{I}),
$$
$$
\nabla \phi_3(\mathbf{z}) = 2\sum_i \frac{1}{\delta - |\mathbf{e}_i^T\mathbf{\Theta z}|^2}\mathbf{\Theta}^H \mathbf{e}_i\mathbf{e}_i^T\mathbf{\Theta z},
$$
$$
\nabla^2 \phi_3(\mathbf{z}) = 2\sum_i \frac{\delta + |\mathbf{e}_i^T\mathbf{\Theta z}|^2}{(\delta - |\mathbf{e}_i^T\mathbf{\Theta z}|^2)^2}\mathbf{\Theta}^H \mathbf{e}_i\mathbf{e}_i^T\mathbf{\Theta}.
$$
$$
\text{(C.1)}
$$

Note

This manuscript was submitted in part to Globecom 2005 and CAMSAP 2005.

REFERENCES

[1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998.

[2] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 684–702, 2003.

[3] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[4] A. van Zelst, "Space division multiplexing algorithms," in *Proceedings of the Mediterranean Electrotechnical Conference (MELECON '00)*, vol. 3, pp. 1218–1221, Lemesos, Cyprus, May 2000.

[5] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, 1998.

[6] A. F. Naguib, N. Seshádri, and A. R. Calderbank, "Increasing data rate over wireless channels: space time coding and signal processing for high data rate wireless communications," *IEEE Signal Processing Magazine*, vol. 17, no. 3, pp. 76–92, 2000.

[7] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1456–1467, 1999.

[8] R. van Nee and R. Prasad, *OFDM for Wireless Multimedia Communications*, Artech House, Boston, Mass, USA, 1st edition, 2000.

[9] L. Deneire, P. Vandenameele, L. van der Perre, B. Gyselinckx, and M. Engels, "A low-complexity ML channel estimator for OFDM," *IEEE Transactions on Communications*, vol. 51, no. 2, pp. 135–140, 2003.

[10] T.-L. Tung, K. Yao, and R. E. Hudson, "Channel estimation and adaptive power allocation for performanceand capacity improvement of multiple-antenna OFDM systems," in *Proceedings of the 3rd IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC '01)*, pp. 82–85, Taiwan, China, March 2001.

[11] Y. Li, N. Seshadri, and S. Ariyavisitakul, "Channel estimation for OFDM systems with transmitter diversity in mobile wireless channels," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 3, pp. 461–471, 1999.

[12] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Transactions on Signal Processing*, vol. 51, no. 6, pp. 1615–1624, 2003.

[13] J. P. Coon, M. Beach, and J. McGeehan, "Optimal training sequences for channel estimation in cyclic-prefix-based single-carrier systems with transmit diversity," *IEEE Signal Processing Letters*, vol. 11, no. 9, pp. 729–732, 2004.

[14] X. Ma, L. Yang, and G. B. Giannakis, "Optimal training for MIMO frequency-selective fading channels," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 453–466, 2005.

[15] H. Minn and N. Al-Dhahir, "Optimal training signals for MIMO OFDM channel estimation," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '04)*, vol. 1, pp. 219–224, Dallas, Tex, USA, November-December 2004.

[16] Q. Sun, D. C. Cox, H. C. Huang, and A. Lozano, "Estimation of continuous flat fading MIMO channels," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 549–553, 2002.

[17] G. Leus and M. Moonen, "Semi-blind channel estimation for block transmissions with non-zero padding," in *Proceedings of the 35th Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 762–766, Pacific Grove, Calif, USA, November 2001.

[18] O. Rousseaux, G. Leus, P. Stoica, and M. Moonen, "A stochastic method for training based channel identification," in *Proceedings of the 7th International Symposium on Signal Processing and Its Applications (ISSPA '03)*, Paris, France, July 2003.

[19] E. G. Larsson and J. Li, "Preamble design for multiple-antenna OFDM-based WLANs with null subcarriers," *IEEE Signal Processing Letters*, vol. 8, no. 11, pp. 285–288, 2001.

[20] 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 Transactions on Communications*, vol. 47, no. 2, pp. 230–238, 1999.

[21] W. Chen and U. Mitra, "Training sequence optimization: comparisons and an alternative criterion," *IEEE Transactions on Communications*, vol. 48, no. 12, pp. 1987–1991, 2000.

[22] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.

[23] M. Chiang, "To layer or not to layer: balancing transport and physical layers in wireless multihop networks," in *Proceedings of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '04)*, vol. 4, pp. 2525–2536, Hong Kong, March 2004.

[24] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: a unified framework for convex optimization," *IEEE Transactions on Signal Processing*, vol. 51, no. 9, pp. 2381–2401, 2003.

[25] J. R. Santos, A. T. Lora, A. G. Expósito, and J. L. M. Ramos, "Finding improved local minima of power system optimization problems by interior-point methods," *IEEE Transactions on Power Systems*, vol. 18, no. 1, pp. 238–244, 2003.

[26] L. Deneire, B. Gyselinckx, and M. Engels, "Training sequence versus cyclic prefix—a new look on single carrier communication," *IEEE Communications Letters*, vol. 5, no. 7, pp. 292–294, 2001.

[27] D. Falconer, S. L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, "Frequency domain equalization for single-carrier broadband wireless systems," *IEEE Communications Magazine*, vol. 40, no. 4, pp. 58–66, 2002.

[28] H. Witschnig, T. Mayer, A. Springer, et al., "A different look on cyclic prefix for SC/FDE," in *Proceedings of the 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '02)*, vol. 2, pp. 824–828, Lisbon, Portugal, September 2002.

[29] O. Rousseaux, G. Leus, and M. Moonen, "A suboptimal iterative method for maximum-likelihood sequence estimation in a multipath context," *EURASIP Journal on Applied Signal Processing*, vol. 2002, no. 12, pp. 1437–1447, 2002.

[30] M. Huemer, H. Witschnig, and J. Hausner, "Unique word based phase tracking algorithms for SC/FDE-systems," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '03)*, vol. 1, pp. 70–74, San Francisco, Calif, USA, December 2003.

[31] Y. Li and L. Yang, "Semi-blind MIMO channel identification based on error adjustment," in *Proceedings of the International Conference on Neural Networks and Signal Processing*, vol. 2, pp. 1429–1432, Nanjing, China, December 2003.

[32] J. P. Coon, M. Sandell, M. Beach, and J. McGeehan, "Channel and noise variance estimation and tracking algorithms for unique-word based single-carrier systems," *IEEE Transactions on Wireless Communications*, vol. 5, no. 6, pp. 1488–1496, 2006.

[33] J. P. Coon and M. Sandell, "Near-optimal unique word design in single-carrier block transmission systems," in *Proceedings of the 8th International Symposium on Signal Processing and Its Applications*, vol. 1, pp. 303–306, Sydney, Australia, August 2005.

[34] D. Hu, L. Yang, Y. Shi, and L. He, "Optimal pilot sequence design for channel estimation in MIMO OFDM systems," *IEEE Communications Letters*, vol. 10, no. 1, pp. 1–3, 2006.

[35] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, USA, 2000.

[36] J. Tellado, *Multicarrier Modulation with Low PAR: Applications to DSL and Wireless*, Kluwer Academic, Norwell, Mass, USA, 2000.

[37] J. G. Proakis, *Digital Communications*, McGraw-Hill, New York, NY, USA, 4th edition, 2001.

[38] V. Erceg, L. Schumacher, P. Kyritsi, et al., "TGn channel models," IEEE 802.11 - 03/940r4, May 2004, http://www.802-wirelessworld.com/.

[39] *Supplement to IEEE standard for information technology - telecommunications and information exchange between systems - local and metropolitan area networks - specific requirements— Part 11: wireless LAN Medium Access Control (MAC) and Physical Layer (PHY)*, IEEE Std 802.11a, 1999.

Justin P. Coon received the B.S. degree (with distinction) in electrical engineering from the Calhoun Honors College at Clemson University, South Carolina, in December 2000, and the Ph.D. degree in wireless communications from the University of Bristol, England, in 2005. He is currently a Senior Research Engineer with Toshiba Research Europe's Telecommunications Research Laboratory in Bristol. His current research interests include optimization, multiple-antenna systems, and ultra-wideband systems. He is a Member of the IEEE.

Magnus Sandell received the M.S. degree in electrical engineering and the Ph.D. degree in signal processing from Luleå University of Technology, Sweden, in 1990 and 1996, respectively. He spent six months as a research assistant with the Division of Signal Processing at the same university before joining Bell Laboratories, Lucent Technologies, Swindon, UK, in 1997. In 2002, he joined Toshiba Research Europe Limited, Bristol, UK, where he is working as a Chief Research Fellow. His research interests include signal processing and digital communications theory. Currently, his focus is on multiple-antenna systems and space-time decoding. He is a Member of the IEEE.