Training Sequence Optimization in MIMO Systems With Colored Interference

Tan F. Wong, *Senior Member, IEEE,* and Beomjin Park

***Abstract*—In this paper, we address the problems of channel estimation and optimal training sequence design for multiple-input multiple-output systems over flat fading channels in the presence of colored interference. In practice, knowledge of the unknown channel is often obtained by sending known training symbols to the receiver. During the training period, we obtain the best linear unbiased estimates of the channel parameters based on the received training block. We determine the optimal training sequence set that minimizes the mean square error of the channel estimator under a total transmit power constraint. In order to obtain the advantage of the optimal training sequence design, long-term statistics of the interference correlation are needed at the transmitter. Hence, this information needs to be estimated at the receiver and fed back to the transmitter. Obviously it is desirable that only a minimal amount of information needs to be fed back from the receiver to gain the advantage in reducing the estimation error of the short-term channel fading parameters. We develop such a feedback strategy in this paper.**

***Index Terms*—Best linear unbiased estimator (BLUE), feedback design, multiple-input multiple-output (MIMO) channel estimation, training sequences.**

I. Introduction

RECENTLY, wireless communication systems using multiple antennas, usually referred to as multiple-input multiple-output (MIMO) systems, have drawn considerable attention, because MIMO systems promise higher capacity [1], [2] than single-antenna systems over fading channels. Different space–time coding techniques [3]–[6] have been proposed to practically achieve the capacity advantages of MIMO systems. To be able to achieve the coding advantage, it is required, for many space–time coding schemes, to obtain accurate channel information at the receiver. In practice, it is common that the unknown channel parameters are estimated by sending known training symbols to the receiver. This training-based channel estimation approach at the receiver is suitable for quasi-static or slowly varying fading channels.

Much work has been done to design training sequences for channel estimation. There have been two major approaches to designing optimal training sequences for both single-antenna systems [7]–[12] and multiple-antenna systems [13]–[18]. One approach is to find training sequences that minimize the channel estimation error [7]–[11], [14], [16] and the other approach is to maximize a lower bound of the channel capacity [15]. A more recent paper [19] presents optimal training sequence designs under these two approaches for correlated MIMO channels. Most of these works assume the presence of white noise. Not much consideration has been given to the case of colored interference, except in [20], where we address the training sequence optimization problem in the presence of a single interferer.

In this paper, we extend our result in [20] to the case in which the colored interference is composed of thermal noise and interference signals transmitted by multiple interferers. We employ the best linear unbiased (BLUE) channel estimator to estimate the channel matrix during the training period. The mean squared error (MSE) of the BLUE channel estimator is used as a performance metric for selecting the training sequence set. We show that the interference covariance matrix decomposes into a Kronecker product of temporal and spatial correlation matrices and that only the temporal correlation needs to be considered in obtaining the optimal training sequence set. The memory of the colored interference induces a nontrivial eigenstructure of the temporal correlation matrix in that some subspaces are less contaminated by the interference. This motivates the problem of judiciously allocating training power to these subspaces. Based on this observation, we determine the optimal training sequence set that minimizes the MSE under a total transmit power constraint. We note that the optimization problem treated in [19] turns out to be similar to the one treated in this paper. In order to obtain the advantage of the optimal training sequence design, we develop an information feedback scheme that requires a minimal amount of information to be fed back from the receiver to approximately obtain the optimal training sequence set at the transmitter.

Numerical results show that we can reduce the MSE of the BLUE channel estimator significantly by using the optimal training sequence set instead of a usual orthogonal training sequence set. We can also achieve comparable estimation performance with the approximate optimal training sequence set obtained by the proposed feedback scheme.

The rest of this paper is organized as follows. In Section II, we describe the MIMO system model and the BLUE channel estimator for the channel matrix based on the received training sequence block. In Section III, the training sequence optimization problem is considered and its solution is given. We also develop the feedback scheme to approximately obtain the optimal training sequence set in Section III. Numerical examples for the

Paper approved by C. Tellambura, the Editor for Modulation and Signal Design of the IEEE Communications Society. Manuscript received May 22, 2003; revised March 19, 2004, and May 31, 2004. This work was supported in part by the Office of Naval Research under Grant N000140210554 and in part by the National Science Foundation under Grant ANI-0020287. This paper was presented in part at the 2004 IEEE GlobeCom Conference.

The authors are with the Wireless Information Networking Group, University of Florida, Gainesville, FL 32611-6130 USA (e-mail: twong@ece.ufl.edu; beomp@ufl.edu).

Digital Object Identifier 10.1109/TCOMM.2004.836558

cases of autoregressive jammers and cochannel interferers are provided in Section IV, and conclusions are drawn in Section V.

II. BLUE Channel Estimator

We consider a transmitter–receiver pair with n_t transmit antennas and n_r receive antennas over a frequency flat fading channel in the presence of colored interference and white thermal noise. We assume that the interference is composed of signals transmitted by M interferers. The jth interferer has n_j transmit antennas. We assume that the transmission from the transmitter to the receiver is packetized. Each packet contains a training frame that is composed of a set of known training sequences, each of which is sent out by a transmit antenna. In matrix notation, the observed training symbols at the receiver for a packet are given by

$$\mathbf{Y} = \mathbf{HS} + \underbrace{\sum_{i=1}^{M} \mathbf{H}_i\mathbf{S}_i + \mathbf{W}}_{\mathbf{E}} = \mathbf{HS} + \mathbf{E} \tag{1}$$

where $\mathbf{S}$ is the $n_t \times N$ transmitted training symbol matrix that is known to the receiver, N is the number of training symbols per transmit antenna, and $\mathbf{S}_i$ is the $n_i \times N$ interference signal matrix from the ith interferer. We assume that symbols in $\mathbf{S}_i$ are zero-mean, complex random variables, correlated across both space and time. In addition, the interference processes are assumed to be wide-sense stationary. We assume that the number of training symbols N is larger than n_t. The $n_r \times n_t$ matrix $\mathbf{H}$ and $n_r \times n_i$ matrix $\mathbf{H}_i$ are the channel matrices from the transmitter and the ith interferer to the receiver, respectively. We assume that the elements in $\mathbf{H}$ and $\mathbf{H}_i$ are independent, identically distributed (i.i.d.) zero-mean, circular-symmetric, complex Gaussian random variables with variance σ^2 and σ_i^2, respectively. In addition, $\mathbf{W}$ is an additive white Gaussian noise (AWGN) matrix and the elements in $\mathbf{W}$ are assumed to be independent, zero-mean, circular-symmetric, complex Gaussian random variables with variance σ_w^2. Finally, $\mathbf{S}_1, \ldots, \mathbf{S}_M$, $\mathbf{H}$, $\mathbf{H}_1, \ldots, \mathbf{H}_M$, and $\mathbf{W}$ are all independent of one another.

Let $\mathbf{y} = \text{vec}(\mathbf{Y})$, $\mathbf{h} = \text{vec}(\mathbf{H})$, and $\mathbf{e} = \text{vec}(\mathbf{E})$, where $\text{vec}(\mathbf{X})$ is the vector obtained by stacking the columns of $\mathbf{X}$ on top of each other [21]. Taking transpose and then vectorizing on both sides of (1), we have

$$\mathbf{y} = \left(\mathbf{S}^T \otimes \mathbf{I}_{n_r}\right)\mathbf{h} + \mathbf{e} \tag{2}$$

where $\otimes$ and $\mathbf{I}_{n_r}$ denote the Kronecker product and $n_r \times n_r$ identity matrix, respectively. In (2), $\mathbf{h}$ is the channel vector with zero mean and covariance matrix $\sigma^2\mathbf{I}_{n_r n_t}$ and $\mathbf{e}$ is the interference-plus-noise vector with zero mean and covariance matrix $\mathbf{Q} = E[\mathbf{e}\mathbf{e}^H]$. From the above, we note that $\mathbf{e}$ is independent of $\mathbf{h}$. During the training period, the BLUE [22] of the channel vector $\mathbf{h}$ based on the received training block $\mathbf{Y}$ can be obtained as

$$\hat{\mathbf{h}} = \left[\left(\mathbf{S}^T \otimes \mathbf{I}_{n_r}\right)^H \mathbf{Q}^{-1}\left(\mathbf{S}^T \otimes \mathbf{I}_{n_r}\right) + \frac{1}{\sigma^2}\mathbf{I}_{n_r n_t}\right]^{-1} \times \left(\mathbf{S}^T \otimes \mathbf{I}_{n_r}\right)^H \mathbf{Q}^{-1}\mathbf{y}. \tag{3}$$

We note that the interference-plus-noise vector is

$$\mathbf{e} = \sum_{i=1}^{M} \text{vec}(\mathbf{H}_i\mathbf{S}_i) + \text{vec}(\mathbf{W}). \tag{4}$$

Let $s_{k,j}^{(i)}$ be the symbol transmitted by the kth antenna of the ith interferer at time j and $R_k^{(i)}(n) = E[s_{k,m}^{(i)} s_{k,m+n}^{(i)*}]$ be the time correlation between the symbols at time instants m and $m+n$ from the kth antenna of the ith interferer. Then it is not hard to show that the $Nn_r \times Nn_r$ noise correlation matrix is given by

$$\mathbf{Q} = \left(\sum_{i=1}^{M} \sigma_i^2 \mathbf{Q}_N^{(i)} + \sigma_w^2 \mathbf{I}_N\right) \otimes \mathbf{I}_{n_r} \tag{5}$$

where

$$\mathbf{Q}_N^{(i)} = \begin{bmatrix} \sum_{k=1}^{n_i} R_k^{(i)}(0) & \cdots & \sum_{k=1}^{n_i} R_k^{(i)}(N-1) \\ \vdots & \ddots & \vdots \\ \sum_{k=1}^{n_i} R_k^{(i)}(N-1) & \cdots & \sum_{k=1}^{n_i} R_k^{(i)}(0) \end{bmatrix}. \tag{6}$$

We note from (5) and (6) that the space correlations between interference symbols from different antennas play no role in the correlation matrix $\mathbf{Q}$. This is due to the i.i.d. assumption we made on the elements of the channel matrices $\mathbf{H}_i$, for $i = 1, \ldots, m$. This turns out to be a crucial property of the interference model, as illustrated below.

The Kronecker product form of $\mathbf{Q}$ in (5) leads to the following simplification of the BLUE for the channel vector $\mathbf{h}$:

$$\hat{\mathbf{h}} = \left\{\left[\left(\mathbf{S}^*\mathbf{A}_N^{-1}\mathbf{S}^T + \frac{1}{\sigma^2}\mathbf{I}_{n_t}\right)^{-1} \mathbf{S}^*\mathbf{A}_N^{-1}\right] \otimes \mathbf{I}_{n_r}\right\}\mathbf{y} \tag{7}$$

where

$$\mathbf{A}_N = \sum_{i=1}^{M} \sigma_i^2 \mathbf{Q}_N^{(i)} + \sigma_w^2 \mathbf{I}_N. \tag{8}$$

Writing (7) back into matrix form, we have

$$\hat{\mathbf{H}} = \mathbf{Y}\mathbf{A}_N^{-1}\mathbf{S}^H\left(\mathbf{S}\mathbf{A}_N^{-1}\mathbf{S}^H + \frac{1}{\sigma^2}\mathbf{I}_{n_t}\right)^{-1}. \tag{9}$$

Moreover, the MSE of the BLUE for $\hat{\mathbf{H}}$ is given by

$$\begin{aligned} \text{MSE}^{(N)} &= \text{tr}\left[\left(\mathbf{S}^T \otimes \mathbf{I}_{n_r}\right)^H \mathbf{Q}^{-1}\left(\mathbf{S}^T \otimes \mathbf{I}_{n_r}\right) + \frac{1}{\sigma^2}\mathbf{I}_{n_r n_t}\right]^{-1} \\ &= n_r \cdot \text{tr}\left[\mathbf{S}^*\mathbf{A}_N^{-1}\mathbf{S}^T + \frac{1}{\sigma^2}\mathbf{I}_{n_t}\right]^{-1}. \end{aligned} \tag{10}$$

We assume that the channel matrices $\mathbf{H}$ and $\mathbf{H}_i$ for $i = 1, \ldots, M$ are short-term statistics that may change from packet to packet. On the other hand, the interference correlation matrix $\mathbf{Q}$ varies at a rate that is much slower than that of the channel matrices. As a result, it is possible for the receiver to estimate $\mathbf{Q}$ using a number of previous packets and feed back relevant information to the transmitter, which can then make use of this information to select the optimal training sequence set for the estimation of $\mathbf{H}$ during the current packet.

III. Training Sequence Optimization

We note that the MSE of the BLUE channel estimator depends on the choice of the training symbol matrix (training sequence set) $\mathbf{S}$. Hence, it is natural to ask whether there is an optimal set of training sequences that gives the best estimation performance. Moreover, it is conceivable that the optimal training sequence set will depend on the characteristic of the interference. Hence, in order to obtain the advantage of employing the optimal training sequence set, information about the interference has to be measured at the receiver and fed back to the transmitter so that it can construct the optimal sequence set. The obvious questions are what information about the interference we should feed back to the transmitter and whether this feedback design is practical or not. We study these questions in this section.

A. Optimal Training Sequence Set

To have a meaningful formulation of the sequence optimization problem, we need to limit the maximum total transmit power of the transmit antenna array to P. The following discussion provides a constructive method to obtain the optimal training sequence set under this restriction.

Our goal is to minimize the MSE of the BLUE channel estimator by selecting the optimal training sequence set $\mathbf{S}$ with the total energy constraint $\mathrm{tr}(\mathbf{S}\mathbf{S}^H) \leq NP$. Therefore, we can express the training sequence set optimization problem as follows:

$$\min_{\mathbf{S}} \mathrm{tr}\left[\mathbf{S}^*\mathbf{A}_N^{-1}\mathbf{S}^T + \frac{1}{\sigma^2}\mathbf{I}_{n_t}\right]^{-1} \text{ subject to } \mathrm{tr}(\mathbf{S}\mathbf{S}^H) \leq NP. \tag{11}$$

Let $\bar{\mathbf{S}} = \mathbf{S}\mathbf{A}_N^{-1/2}$. We can rewrite the optimization problem in (11) in the following form:

$$\min_{\bar{\mathbf{S}}} \mathrm{tr}\left[\bar{\mathbf{S}}\bar{\mathbf{S}}^H + \frac{1}{\sigma^2}\mathbf{I}_{n_t}\right]^{-1} \text{ subject to } \mathrm{tr}(\bar{\mathbf{S}}\mathbf{A}_N\bar{\mathbf{S}}^H) \leq NP. \tag{12}$$

Further, let $\mu_1, \ldots, \mu_{n_t}$ be the nonnegative eigenvalues of $\bar{\mathbf{S}}\bar{\mathbf{S}}^H$ arranged in a descending order and $\lambda_1^{(N)}, \ldots, \lambda_N^{(N)}$ be the positive eigenvalues of $\mathbf{A}_N$ arranged in an ascending order. To proceed, we need to make use of the following result, whose proof can be found, for example, in [23, pp. 249].

Lemma 1: Suppose that $\mathbf{X}$ and $\mathbf{Y}$ are two Hermitian $N \times N$ matrices. Arrange the eigenvalues $x_1, \ldots, x_N$ of $\mathbf{X}$ in a descending order and the eigenvalues $y_1, \ldots, y_N$ of $\mathbf{Y}$ in an ascending order. Then $\mathrm{tr}(\mathbf{X}\mathbf{Y}) \geq \sum_{i=1}^N x_i y_i$.

Applying this lemma to the constraint in (12), we can bound

$$\mathrm{tr}(\bar{\mathbf{S}}\mathbf{A}_N\bar{\mathbf{S}}^H) = \mathrm{tr}(\mathbf{A}_N\bar{\mathbf{S}}^H\bar{\mathbf{S}}) \geq \sum_{i=1}^{n_t} \mu_i \lambda_i^{(N)}. \tag{13}$$

Now, consider the following relaxed optimization problem:

$$\min_{\mu_1,\ldots,\mu_{n_t}} \sum_{i=1}^{n_t}\left(\mu_i + \frac{1}{\sigma^2}\right)^{-1}$$
$$\text{subject to} \sum_{i=1}^{n_t} \mu_i\lambda_i^{(N)} \leq NP \text{ and } \mu_1 \geq \cdots \geq \mu_{n_t} \geq 0. \tag{14}$$

This relaxed optimization problem can be solved by the standard Karush–Kuhn–Tucker condition technique [25] since the cost function and the constraint are both convex. Indeed, let

$$n_* = \max\left\{k \in \{1,2,\ldots,n_t\} : \sqrt{\lambda_k^{(N)}} \cdot \sum_{i=1}^{k}\sqrt{\lambda_i^{(N)}} - \sum_{i=1}^{k}\lambda_i^{(N)} < \sigma^2 NP\right\}. \tag{15}$$

We note that n_* above is well defined and that the inequality that defines n_* in (15) holds for $k = 1, \ldots, n_*$, while it does not hold for $k = n_* + 1, \ldots, n_t$. With this definition, it is not hard to show that the optimal solution is given by

$$\mu_k^* = \begin{cases} \frac{1}{\sigma^2}\left(\frac{\sigma^2 NP + \sum_{i=1}^{n_*}\lambda_i^{(N)}}{\sqrt{\lambda_k^{(N)}}\cdot\sum_{i=1}^{n_*}\sqrt{\lambda_i^{(N)}}} - 1\right) & \text{for } k = 1, \ldots, n_* \\ 0 & \text{for } k = n_*+1, \ldots, n_t. \end{cases} \tag{16}$$

We note that this solution has the standard water-filling [26] interpretation.

If we can construct a matrix $\bar{\mathbf{S}}$ such that the eigenvalues of $\bar{\mathbf{S}}\bar{\mathbf{S}}^H$ are exactly the solution of the above relaxed optimization problem and that $\mathrm{tr}(\bar{\mathbf{S}}\mathbf{A}_N\bar{\mathbf{S}}^H) = \sum_{i=1}^{n_t}\mu_i^*\lambda_i^{(N)}$, then this choice of $\bar{\mathbf{S}}$ will be a solution of the original sequence optimization problem in (12). It is easy to see that this can be done and the resulting optimal training sequence set is given by

$$\mathbf{S}_* = \mathbf{V}\,\mathrm{diag}\left[\sqrt{\mu_1^*\lambda_1^{(N)}}, \ldots, \sqrt{\mu_{n_t}^*\lambda_{n_t}^{(N)}}\right]\bar{\mathbf{U}}_N^H \tag{17}$$

where $\mathbf{V}$ is an arbitrary $n_t \times n_t$ unitary matrix and $\bar{\mathbf{U}}_N$ is the $N \times n_t$ matrix whose columns are the eigenvectors of $\mathbf{A}_N$ corresponding to the n_t smallest eigenvalues of $\mathbf{A}_N$. With this optimal choice of training sequence set, the minimum estimation error achieved is given by

$$\mathrm{MSE}_*^{(N)} = \frac{n_r\sigma^2\left(\sum_{i=1}^{n_*}\sqrt{\lambda_i^{(N)}}\right)^2}{\sigma^2 NP + \sum_{i=1}^{n_*}\lambda_i^{(N)}} + n_r(n_t - n_*)\sigma^2. \tag{18}$$

A physical interpretation of this solution is that the optimal training sequence set put its power to where the effect of the interference is the smallest, hence the estimation error can be minimized. We note that the optimal training sequence set is an orthogonal set if $\mathbf{V}$ is chosen to be an identity matrix. However, the optimal training sequence set, in general, is not necessarily orthogonal. For instance, it is possible to obtain a choice of $\mathbf{V}$ which spreads power evenly across the transmit antennas with the use of nonorthogonal sequences. To do so, we need to construct a unitary $\mathbf{V}$ to make the diagonal elements of

$$\mathbf{S}_*\mathbf{S}_*^H = \mathbf{V}\,\mathrm{diag}\left[\mu_1^*\lambda_1^{(N)}, \ldots, \mu_{n_t}^*\lambda_{n_t}^{(N)}\right]\mathbf{V}^H \tag{19}$$

the same. This is shown to be possible in [24], and such a $\mathbf{V}$ can be constructed using a simple iterative procedure.

We note that not only does the choice of optimal sequence set minimize the estimation error, but this choice also simplifies the implementation complexity of the BLUE for $\mathbf{H}$. It is easy to see that the BLUE channel estimator in (9) reduces to

$$\hat{\mathbf{H}} = \mathbf{Y}\bar{\mathbf{U}}_N \mathrm{diag}\left[\omega_1, \ldots, \omega_{n_t}\right] \mathbf{V}^H \tag{20}$$

where

$$\omega_k = \begin{cases} \frac{1+\sigma^2\mu_k^*}{\sqrt{\mu_k^*\lambda_i^{(N)}}}, & \text{for } k = 1, \ldots, n_* \\ 0, & \text{for } k = n_* + 1, \ldots, n_t. \end{cases} \tag{21}$$

Thus, the complexity of the BLUE with the optimal sequence set reduces to $\mathcal{O}(n_t n_r N)$.

B. Feedback Design

In order to obtain the advantage of the optimal training sequence design, long-term statistics of the interference correlation need to be estimated at the receiver and fed back to the transmitter. We note from Section III-A that the optimal training sequence set depends on the channel gain variance σ^2 and the eigenstructure of the matrix $\mathbf{A}_N$ in (8). As a result, only these two long-term statistics need to be estimated at the receiver. Obviously it is desirable that only a minimal amount of information is needed to be fed back from the receiver to gain the advantage in reducing the estimation error of the short-term channel fading matrix. In this section, we develop such a feedback scheme based on the fact that a suitable Toeplitz matrix can be approximated by a circulant matrix.

Since the interferer signals are wide-sense stationary, $\mathbf{A}_N$ takes the form of a Toeplitz matrix. Indeed, consider a sequence of complex numbers $\{a_l\}_{l=-\infty}^{\infty}$ such that $a_l = a_{-l}^*$ and the elements of $\mathbf{A}_N$ at the ith row and jth column is given by a_{i-j}. The sequence $\{a_l\}_{l=-\infty}^{\infty}$ is obtained by sampling the autocorrelation function of the interference at the symbol rate. In addition, if the sequence $\{a_l\}_{l=-\infty}^{\infty}$ is absolutely summable, then it is shown in [27] that the Toeplitz matrix $\mathbf{A}_N$ can be approximated by the circulant matrix

$$\tilde{\mathbf{A}}_N = \mathbf{F}_N \mathbf{\Lambda}_N \mathbf{F}_N^H \tag{22}$$

where $\mathbf{F}_N$ is the $N \times N$ FFT matrix, i.e., the (k, l)th element of $\mathbf{F}_N$ is $(1/\sqrt{N})e^{-j(2\pi(k-1)(l-1)/N)}$, and $\mathbf{\Lambda}$ is an $N \times N$ diagonal matrix with $\delta_1^{(N)}, \ldots, \delta_N^{(N)}$ as its diagonal elements.

A reasonable way [27] to obtain $\delta_l^{(N)}$, for $l = 1, \ldots, N$, is

$$\delta_l^{(N)} = a_0 + 2\Re\left(\sum_{k=1}^{N-1} a_k e^{j\frac{2\pi k(l-1)}{N}}\right). \tag{23}$$

With this choice of $\delta_1^{(N)}, \ldots, \delta_N^{(N)}$, it can be shown that $\tilde{\mathbf{A}}_N$ approaches $\mathbf{A}_N$ as N approaches infinity [27]. Moreover, if we arrange $\delta_1^{(N)}, \ldots, \delta_N^{(N)}$ in an ascending order, we have

$$\lim_{N\to\infty}\left|\delta_{[l]}^{(N)} - \lambda_l^{(N)}\right| = 0 \tag{24}$$

for $l = 1, \ldots, n_t$, and $\delta_{[l]}^{(N)}$ is the lth smallest eigenvalue among the set $\{\delta_l^{(N)}\}_{l=1}^{N}$.

Now we turn to the estimation of σ^2 and $\{a_k\}_{k=0}^{N-1}$. As mentioned before, they are both long-term statistics and hence should be estimated based on the observed training frames of the previous K packets, where K is smaller than the number of packets during which the long-term statistics remain the same. Toward this end, let $s_{i,l}(n)$ denote the (i,l)th element of the training matrix $\mathbf{S}(n)$ and $y_{i,l}(n)$ denote the (i,l)th element of the observed training matrix $\mathbf{Y}(n)$ as defined in (1) for the nth packet, respectively. Then it is not hard to see that, for $i = 1, 2, \ldots, n_r$ and $k = 0, 1, \ldots, N-1$,

$$(N-k)a_k = E\left[\frac{1}{K}\sum_{j=n-K+1}^{n}\sum_{l=1}^{N-k} y_{i,l}(j)y_{i,l+k}^*(j) - \sigma^2 N R_n^s(k)\right] \tag{25}$$

where

$$R_n^s(k) = \frac{1}{NK}\sum_{j=n-K+1}^{n}\sum_{l=1}^{N-k}\sum_{m=1}^{n_t} s_{m,l}(j)s_{m,l+k}^*(j). \tag{26}$$

In addition, let $\mathbf{P}_{\mathbf{S}}(n)$ be the $N \times N$ projection matrix onto the subspace perpendicular to the one spanned by the rows of the training matrix $\mathbf{S}(n)$ for the nth packet. Denote the (i, l)th element of $\mathbf{Y}(n)\mathbf{P}_{\mathbf{S}}(n)$ by $\tilde{y}_{i,l}(n)$ and the (k, l)th element of $\mathbf{P}_{\mathbf{S}}(n)$ by $p_{k,l}(n)$. Then it can be shown that, for $i = 1, 2, \ldots, n_r$, we have

$$a_0 R_n^p(0) + 2\sum_{k=1}^{N-1}\Re\left[a_k R_n^p(k)\right] = E\left[\frac{1}{K}\sum_{j=n-K+1}^{n}\sum_{l=1}^{N}\left|\tilde{y}_{i,l}(j)\right|^2\right] \tag{27}$$

where

$$R_n^p(k) = \frac{1}{K}\sum_{j=n-K+1}^{n}\sum_{l=1}^{N-k} p_{l,l+k}(j). \tag{28}$$

We note that (25) for $k = 0, \ldots, N-1$ together with (27) provides us $2N$ equations to solve for the $2N$ unknowns σ^2, a_0 (both real-valued), and a_k for $k = 1, \ldots, N-1$ (all complex-valued). Thus, estimates of σ^2 and a_k for $k = 0, 1, \ldots, N-1$ can be obtained by solving this set of linear equations with the expectation terms replaced by their usual estimates as follows:

$$\begin{aligned}
&a_0 + \sigma^2 R_n^s(0) \\
&\quad = \frac{1}{n_r N K}\sum_{i=1}^{n_r}\sum_{j=n-K+1}^{n}\sum_{l=1}^{N}\left|y_{i,l}(j)\right|^2 \\
&\quad\vdots \\
&a_k + \sigma^2 R_n^s(k) \\
&\quad = \frac{1}{n_r N K}\sum_{i=1}^{n_r}\sum_{j=n-K+1}^{n}\sum_{l=1}^{N-k} y_{i,l}(j)y_{i,l+k}^*(j) \\
&\quad\vdots \\
&a_{N-1} + \sigma^2 R_n^s(N-1) \\
&\quad = \frac{1}{n_r N K}\sum_{i=1}^{n_r}\sum_{j=n-K+1}^{n} y_{i,1}(j)y_{i,N}^*(j) \\
&a_0 R_n^p(0) + 2\sum_{k=1}^{N-1}\Re\left[a_k R_n^p(k)\right] \\
&\quad = \frac{1}{n_r K}\sum_{i=1}^{n_r}\sum_{j=n-K+1}^{n}\sum_{l=1}^{N}\left|\tilde{y}_{i,l}(j)\right|^2.
\end{aligned} \tag{29}$$

In the above, the biased estimators $((N-k)/N)[(1/n_r)\sum_{i=1}^{n_r}\sum_{j=n-K+1}^{n}\sum_{l=1}^{N-k} y_{i,l}(j)y_{i,l+k}^*(j) - \sigma^2 R_n^s(k)]$, for $k=0,1,\ldots,N-1$, have been employed to approximate the corresponding expectations on the right-hand side of (25). We note that the use of these biased estimators is similar to the use of biased autocorrelation function estimators in the Yule–Walker method of estimating the power spectral density of an AR process [28].

In summary, we can use the solution of (29) to estimate the autocorrelation function of the interference at the receiver and obtain estimates of the $\delta_l^{(N)}$'s by (23). Then the estimated value of σ^2, the n_t smallest $\delta_l^{(N)}$'s, and the corresponding indices are fed back to the transmitter. At the transmitter, we can replace $\bar{\mathbf{U}}_N$ by the columns of $\mathbf{F}_N$ that are indexed by the feedback indices to construct the optimal training sequence set for the current packet.

We note that the estimates of the a_k's obtained by solving (29) do not guarantee the resulting estimates of the δ_k's and σ^2 to be positive, although this is almost always the case when N, K, and n_r are sufficiently large. When the estimate of σ^2 is negative, we heuristically use the absolute value of the estimate instead. In addition, we do not use those δ_k's with negative estimates in finding the n_t minimum values of $\lambda_1^{(N)},\ldots,\lambda_N^{(N)}$ as described before.

C. Asymptotic Estimation Performance Gain

It is illustrative as well as practical to develop a simple measure that can tell us how much advantage we can obtain by employing the optimal training sequence set over other choices of training sequences. For instance, if the receiver determines that there is not much to gain by using the optimal training sequences, it can inform the transmitter to keep on using the current ones. To this end, we employ equal-power orthogonal training sequences as our baseline for comparison, since these training sequences are commonly [13]–[16] suggested when the noise is white.

First, we want to obtain the worst-case MSE, $\mathrm{MSE}_{\max}^{(N)}$, when equal-power orthogonal training sequences are employed and the total transmit power is P. It is not too hard to see that

$$\begin{aligned}\max_{\mathbf{SS}^H=NP\mathbf{I}} \mathrm{tr}\left[\mathbf{SA}_N^{-1}\mathbf{S}^H+\frac{1}{\sigma^2}\mathbf{I}_{n_t}\right]^{-1} &\le \frac{n_t}{\min\limits_{\mathbf{SS}^H=NP\mathbf{I}} \lambda_{\min}\left(\mathbf{SA}_N^{-1}\mathbf{S}^H\right)+\frac{1}{\sigma^2}}\\ &= \frac{n_t}{\frac{NP}{n_t}\cdot\frac{1}{\lambda_N^{(N)}}+\frac{1}{\sigma^2}}\\ \max_{\mathbf{SS}^H=NP\mathbf{I}} \mathrm{tr}\left[\mathbf{SA}_N^{-1}\mathbf{S}^H+\frac{1}{\sigma^2}\mathbf{I}_{n_t}\right]^{-1} &\ge \frac{n_t}{\min\limits_{\mathbf{SS}^H=NP\mathbf{I}} \lambda_{\max}\left(\mathbf{SA}_N^{-1}\mathbf{S}^H\right)+\frac{1}{\sigma^2}}\\ &= \frac{n_t}{\frac{NP}{n_t}\cdot\frac{1}{\lambda_{N-n_t+1}^{(N)}}+\frac{1}{\sigma^2}}\end{aligned} \tag{30}$$

where $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ are the minimum and maximum eigenvalues of a Hermitian matrix, respectively, and $\lambda_1^{(1)},\ldots,\lambda_N^{(N)}$ are the eigenvalues of $\mathbf{A}_N$ arranged in an ascending order. From (10), we can bound the worst-case MSE as

$$\frac{n_r n_t}{\frac{NP}{n_t}\cdot\frac{1}{\lambda_{N-n_t+1}^{(N)}}+\frac{1}{\sigma^2}} \le \mathrm{MSE}_{\max}^{(N)} \le \frac{n_r n_t}{\frac{NP}{n_t}\cdot\frac{1}{\lambda_N^{(N)}}+\frac{1}{\sigma^2}}. \tag{31}$$

On the other hand, from (18), when the optimal training sequence set is employed, the minimum MSE can be bounded by

$$\frac{n_r n_*}{\frac{NP}{n_*}\frac{1}{\lambda_1^{(N)}}+\frac{1}{\sigma^2}\frac{\lambda_{n_*}^{(N)}}{\lambda_1^{(N)}}} \le \mathrm{MSE}_*^{(N)} \le \frac{n_r n_*}{\frac{NP}{n_*}\frac{1}{\lambda_{n_*}^{(N)}}+\frac{1}{\sigma^2}\frac{\lambda_1^{(N)}}{\lambda_{n_*}^{(N)}}}. \tag{32}$$

Combining (31) and (32), we can bound the ratio between the minimum MSE and the worst-case MSE by

$$\frac{n_*}{n_t}\cdot\frac{\frac{1}{\lambda_N^{(N)}}+\frac{1}{\sigma^2}\frac{n_t}{NP}}{\frac{1}{\lambda_1^{(N)}}+\frac{1}{\sigma^2}\frac{\lambda_{n_*}^{(N)}}{\lambda_1^{(N)}}\frac{n_t}{NP}} \le \frac{\mathrm{MSE}_*^{(N)}}{\mathrm{MSE}_{\max}^{(N)}} \le \frac{n_*}{n_t}\cdot\frac{\frac{1}{\lambda_{N-n_t+1}^{(N)}}+\frac{1}{\sigma^2}\frac{n_t}{NP}}{\frac{1}{\lambda_{n_*}^{(N)}}+\frac{1}{\sigma^2}\frac{\lambda_1^{(N)}}{\lambda_{n_*}^{(N)}}\frac{n_t}{NP}}. \tag{33}$$

This MSE ratio gives the maximum possible relative reduction in the estimation error that we can obtain by using the optimal sequence set under a specific set of interferers. Here we obtain a simpler performance metric by considering the asymptotic value of this MSE ratio when N is very large.

To do so, we employ the following results regarding the extremal eigenvalues of the sequence of Toeplitz matrices $\{\mathbf{A}_N\}$ in [29, Ch. 5]. Suppose that the sampled autocorrelation sequence, $\{a_n\}_{n=-\infty}^{\infty}$, of the wide-sense stationary interference process is absolutely summable. Let

$$A(\omega)=\sum_{n=-\infty}^{\infty} a_n e^{-jn\omega}$$

be the discrete-time Fourier transform of $\{a_n\}_{n=-\infty}^{\infty}$. Then, for $i=1,2,\ldots,n_t$, we have

$$\begin{aligned}\lim_{N\to\infty}\lambda_i^{(N)} &= \min_{0\le\omega\le 2\pi} A(\omega)\\ \lim_{N\to\infty}\lambda_{N-i+1}^{(N)} &= \max_{0\le\omega\le 2\pi} A(\omega).\end{aligned} \tag{34}$$

From (15) and (34), we see that $\lim_{N\to\infty} n_* = n_t$. Applying this and (34) to (33), the asymptotic maximum MSE reduction ratio is given as follows:

$$\Gamma=\lim_{N\to\infty}\frac{\mathrm{MSE}_*^{(N)}}{\mathrm{MSE}_{\max}^{(N)}}=\frac{\min\limits_{0\le\omega\le 2\pi} A(\omega)}{\max\limits_{0\le\omega\le 2\pi} A(\omega)}. \tag{35}$$

IV. Numerical Examples

In this section, we consider two examples to illustrate the potential advantage of employing the optimal training sequence set. The first example considers the case when the interference signals are described by first-order autoregressive (AR) random processes. The second example considers the case in which the interference is caused by cochannel interferers whose signal structures are exactly the same as that of the desired signal. We assume that the desired user has two transmit antennas and three receive antennas. In each example, we evaluate the MSEs of the BLUE channel estimator when the following three different training sequence sets are employed:

1) the Hadamard sequence set, i.e., the first two rows of a Hadamard matrix are used as the training sequences;
2) the optimal training sequence set described in Section III-A; and
3) the approximate optimal training sequence set described in Section III-B.

A. AR(1) Jammers

We assume that there are two jammers in the system. Both jammers have one transmit antenna. The interference signals from the jammers are modeled by two first-order AR processes with AR parameters $0 < \alpha_1, \alpha_2 < 1$, respectively. For instance, the AR model of the first jammer is given by

$$s_t^{(1)} = \alpha_1 s_{t-1}^{(1)} + u_{1,t} \tag{36}$$

where $u_{1,t}$ is a white Gaussian random process with zero mean and variance $\sigma_{u,1}^2$. The AR parameter can be interpreted as the intensity of correlation among the symbols of the jammer. It is easy to verify that, for this case, we have

$$A(\omega) = \sum_{m=1}^{2} \frac{\sigma_m^2 \sigma_{u,m}^2}{1 + \alpha_m^2 - 2\alpha_m \cos\omega} + \sigma_w^2. \tag{37}$$

Hence, the asymptotic maximum MSE reduction ratio is given by

$$\Gamma = \frac{\sum_{m=1}^{2} \frac{\sigma_m^2 P_m}{\sigma_w^2} \dfrac{1-\alpha_m}{1+\alpha_m} + 1}{\sum_{m=1}^{2} \frac{\sigma_m^2 P_m}{\sigma_w^2} \dfrac{1+\alpha_m}{1-\alpha_m} + 1} \tag{38}$$

where $P_m = \sigma_{u,m}^2/(1-\alpha_m^2)$ is the transmit power of the mth jammer.

The MSEs of the BLUE channel estimator with the three different training sequence sets are shown and compared in Figs. 1 and 2 for the two different combinations of α_1 and α_2. In each case, we consider different lengths ($N = 16$, 32, 64, 128, 256, 512, and 1024) of the training sequences and different received signal-to-interference ratios ($(\sigma^2 P/\sigma_i^2 P_i) = 0$ dB and -20 dB for $i = 1$ and 2). The received signal-to-noise ratio (SNR) $(\sigma^2 P/\sigma_w^2)$ is set to 10 dB. The technique described in Section III-B is employed to obtain the approximate optimal training sequences. We have assumed that the received training signals from ten previous packets are employed to estimate the jammer information at the receiver. The training sequences used in the previous ten packets are the Hadamard sequences described above.

From these figures, we observe that there exist only minimal differences between the MSEs of using the optimal training sequence set and approximate optimal training sequence set. Obviously, this is a desirable result, because it indicates that we can obtain comparable performance to the optimal training sequence set by estimating the jammer information at the receiver and feeding back only a small amount of information to the transmitter. In general, we see that the optimal training sequence set significantly outperforms the Hadamard sequence set in all the cases considered. The advantage of using the optimal

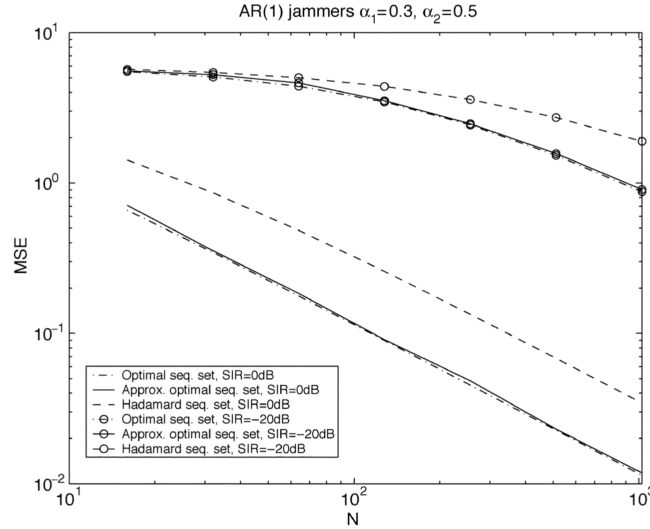


Fig. 1. Comparison of MSEs obtained by using different training sequence sets. Two AR(1) jammers with $\alpha_1 = 0.3$ and $\alpha_2 = 0.5$.

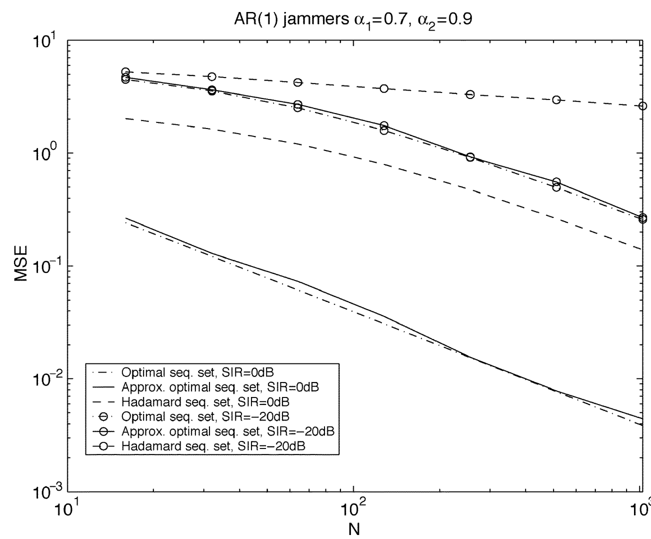


Fig. 2. Comparison of MSEs obtained by using different training sequence sets. Two AR(1) jammers with $\alpha_1 = 0.7$ and $\alpha_2 = 0.9$.

training sequence set increases as the correlation parameters α_i's increase. The asymptotic maximum MSE reduction ratios for the cases considered above are shown in Table I. For comparison, the MSE reduction ratios obtained by using the optimal sequenceset against the Hadamard sequence set for $N = 1024$ are also included in Table I. We can deduce from the table that the Hadamard sequence set is rather inefficient. In addition, much more reduction in MSE can be obtained using the optimal sequence set when both of the α_i's are close to 1.

B. Cochannel Interferers

In this example, we assume that the interference is caused by two cochannel interferers whose signal format is similar to that of the desired user. More precisely, let us assume that the transmitted signal at the ith transmit antenna of the mth interferer is given by

$$s_i^{(m)}(t) = \sqrt{\frac{P_m}{2}} \sum_{l=-\infty}^{\infty} b_{i,l}^{(m)} \psi(t - lT - \tau_m) \tag{39}$$

TABLE I
COMPARISON OF ASYMPTOTIC MAXIMUM MSE REDUCTION RATIO AND MSE RATIO BETWEEN USING OPTIMAL AND HADAMARD SEQUENCES IN THE CASE OF AR JAMMERS

$\alpha_1 = 0.3,\ \alpha_2 = 0.5$		
SIR	Asymp. max. MSE reduct. ratio	$\frac{\text{MSE with optimal seqs.}}{\text{MSE with Hadamard seqs.}}$ $(N = 1024)$
0dB	−7.08dB	−4.81dB
−20dB	−7.46dB	−3.37dB

$\alpha_1 = 0.7,\ \alpha_2 = 0.9$		
SIR	Asymp. max. MSE reduct. ratio	$\frac{\text{MSE with optimal seqs.}}{\text{MSE with Hadamard seqs.}}$ $(N = 1024)$
0dB	−18.77dB	−15.56dB
−20dB	−20.30dB	−10.05dB

where $b_{i,l}^{(m)}$ is the sequence of data symbols, which are assumed to be i.i.d. binary random variables with zero mean and unit variance, from the ith antenna of the mth interferer, $\psi(t)$ is the symbol waveform, T is the symbol interval, and τ_m is the symbol timing difference between the mth interferer and the desired signal. Without loss of generality, we can assume that $\tau_m \in [0, T)$. We also assume that $\int_{-\infty}^{\infty} |\psi(t)|^2 dt = 1$.

With the model described above, the elements of the interference signal matrix $\mathbf{S}_m$ in (1) are samples at the matched filter output at the receiver at time kT. Specifically, the (i, k)th element of $\mathbf{S}_m$ is given by

$$s_{i,k}^{(m)} = \sum_{l=-\infty}^{\infty} b_{i,l}^{(m)} \hat{\psi}\left((k-l)T - \tau_m\right) \tag{40}$$

where

$$\hat{\psi}(t) = \int_{-\infty}^{\infty} \psi(t-s)\psi^*(s)ds \tag{41}$$

is the autocorrelation of the symbol waveform. Thus, it is easy to see that the sampled autocorrelation sequence

$$a_n = \sum_{m=1}^{2} \sigma_m^2 P_m \sum_{l=-\infty}^{\infty} \hat{\psi}\left((l-n)T - \tau_m\right) \hat{\psi}^*(lT - \tau) + \sigma_w^2 \delta_n \tag{42}$$

and its discrete-time Fourier transform is given by

$$\begin{aligned} A(\omega) &= \sum_{m-1}^{2} \sigma_m^2 P_m \left| \sum_{n=-\infty}^{\infty} \hat{\psi}(nT + \tau_m) e^{-j\omega n} \right|^2 + \sigma_w^2 \\ &= \sum_{m-1}^{2} \frac{\sigma_m^2 P_m}{T^2} \left| \sum_{n=-\infty}^{\infty} \hat{\Psi}\left(\frac{\omega - 2\pi k}{T}\right) e^{\frac{j(\omega - 2\pi k)\tau_m}{T}} \right|^2 \\ &\quad + \sigma_w^2 \end{aligned} \tag{43}$$

where P_m is the transmit power from the mth interferer, $\hat{\Psi}(\Omega) = |\Psi(\Omega)|^2$, and $\Psi(\Omega)$ is the Fourier transform of the symbol waveform $\psi(t)$.

To illustrate how the use of the optimal training sequence set can benefit the channel estimation process, let us consider the following two common symbol waveforms:

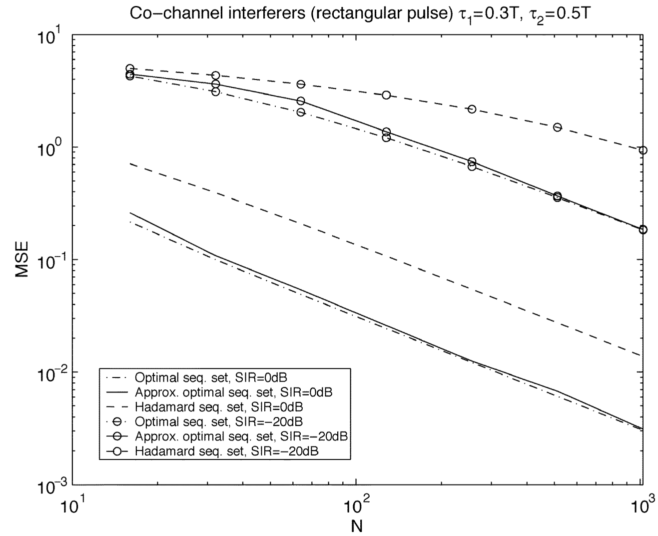


Fig. 3. Comparison of MSEs obtained by using different training sequence sets. Two cochannel interferers with rectangular waveforms and delays $\tau_1 = 0.3T$, $\tau_2 = 0.5T$.

1) Rectangular Symbol Waveform: In this case

$$\psi(t) = \begin{cases} \frac{1}{T}, & 0 \le t \le T \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad \hat{\psi}(t) = \begin{cases} \frac{t}{T}, & 0 \le t \le T \\ 2 - \frac{t}{T}, & T \le t \le 2T \\ 0, & \text{otherwise.} \end{cases}$$

From (43), we have

$$\begin{aligned} A(\omega) = \sum_{m=1}^{2} \sigma_m^2 P_m \Bigg[\left(\frac{\tau_m}{T}\right)^2 &+ \left(1 - \frac{\tau_m}{T}\right)^2 \\ &+ 2\left(\frac{\tau_m}{T}\right)\left(1 - \frac{\tau_m}{T}\right)\cos\omega \Bigg] + \sigma_w^2. \end{aligned} \tag{44}$$

Hence, the asymptotic maximum MSE reduction ratio is

$$\Gamma = \frac{\sum_{m=1}^{2} \frac{\sigma_m^2 P_m}{\sigma_w^2} \left(2\frac{\tau_m}{T} - 1\right)^2 + 1}{\sum_{m=1}^{2} \frac{\sigma_m^2 P_m}{\sigma_w^2} + 1}. \tag{45}$$

From (45), the use of the optimal training sequence provides no gain when the cochannel interferers are symbol-synchronous to the desired user signal, i.e., $\tau_1 = \tau_2 = 0$. On the other hand, when $\tau_1 = \tau_2 = 0.5T$, the asymptotic maximum MSE reduction ratio attains its smallest possible value. This means that we can almost completely eliminate the effect of the interferers by using the set of long optimal training sequences.

As before, we compare the MSEs of the BLUE channel estimator with the three different training sequence sets in Fig. 3 by considering the case in which $\tau_1 = 0.3T$ and $\tau_2 = 0.5T$. The other parameters are chosen as in the AR jammer example before. Again, from Fig. 3, we observe that there exist only minimal differences between the MSEs of using the optimal training sequence set and approximate optimal training sequence set, and that the optimal training sequence set significantly outperforms the Hadamard sequence set in all the cases considered. The asymptotic maximum MSE reduction ratios for the cases considered above are shown in Table II. For comparison, the MSE reduction ratios obtained by using the optimal sequence set against the Hadamard sequence set for $N = 1024$ are also included in Table II. We can deduce from the table that the Hadamard sequence set is rather inefficient.

TABLE II
COMPARISON OF ASYMPTOTIC MAXIMUM MSE REDUCTION RATIO AND MSE RATIO BETWEEN USING OPTIMAL AND HADAMARD SEQUENCES IN THE CASE OF COCHANNEL INTERFERERS

Rectangular waveform $\tau_1 = 0.3T$, $\tau_2 = 0.5T$		
SIR	Asymp. max. MSE reduct. ratio	$\frac{\text{MSE with optimal seqs.}}{\text{MSE with Hadamard seqs.}}$ $(N = 1024)$
0dB	-9.07dB	-6.55dB
-20dB	-10.94dB	-7.08dB

ISI-free waveform $\tau_1 = 0.3T$, $\tau_2 = 0.5T$		
SIR	Asymp. max. MSE reduct. ratio	$\frac{\text{MSE with optimal seqs.}}{\text{MSE with Hadamard seqs.}}$ $(N = 1024)$
0dB	-6.73dB	-4.54dB
-20dB	-7.62dB	-4.34dB

2) ISI-Free Symbol Waveform With Raised Cosine Spectrum [30]: In this case, we have

$$\hat{\Psi}(\Omega) = \begin{cases} T, & 0 \leq |\Omega| \leq \frac{\pi(1-\beta)}{T} \\ \frac{T}{2}\left\{1+\cos\left[\frac{T}{2\beta}\left(|\Omega| - \frac{\pi(1-\beta)}{T}\right)\right]\right\}, & \frac{\pi(1-\beta)}{T} \leq |\Omega| \leq \frac{\pi(1+\beta)}{T} \\ 0, & |\Omega| \geq \frac{\pi(1+\beta)}{T} \end{cases}$$

where $0 < \beta \leq 1$ is the roll-off factor. Since $\sum_{k=-\infty}^{\infty} \hat{\Psi}((\omega - 2\pi k)/T) = T$ for all ω and $\hat{\Psi}(\Omega)$ is positive, it can be deduced from (43) that $\max_{0 \leq \omega \leq 2\pi} A(\omega) = \sum_{m=1}^{2} \sigma_m^2 P_m + \sigma_w^2$. To find $\min_{0 \leq \omega \leq 2\pi} A(\omega)$, because of symmetry of $\hat{\Psi}(\Omega)$, it is enough to consider the interval $\omega \in [\pi(1-\beta), \pi]$. Over this interval, by (43), we have

$$\begin{aligned} A(\omega) = \sum_{m=1}^{2} \frac{\sigma_m^2 P_m}{4} & \times \left\{ \left[1+\cos\left[\frac{\omega - \pi(1-\beta)}{2\beta}\right]\right]^2 \right. \\ & + \left[1+\cos\left[\frac{\omega - \pi(1+\beta)}{2\beta}\right]\right]^2 + 2\cos\left(\frac{2\pi\tau_m}{T}\right) \\ & \times \left[1+\cos\left[\frac{\omega - \pi(1-\beta)}{2\beta}\right]\right] \\ & \left. \times \left[1+\cos\left[\frac{\omega - \pi(1+\beta)}{2\beta}\right]\right] \right\} + \sigma_w^2. \end{aligned} \quad (46)$$

Simple calculus reveals that $\min_{0 \leq \omega \leq 2\pi} A(\omega) = \sum_{m=1}^{2} \sigma_m^2 P_m \cos^2(\pi\tau_m/T) + \sigma_m^2$. Thus, the asymptotic maximum MSE reduction ratio is

$$\Gamma = \frac{\sum_{m=1}^{2} \frac{\sigma_m^2 P_m}{\sigma_w^2} \cos^2\left(\frac{\pi\tau_m}{T}\right) + 1}{\sum_{m=1}^{2} \frac{\sigma_m^2 P_m}{\sigma_w^2} + 1}. \quad (47)$$

From (47), the use of the optimal training sequence provides no gain when the co-channel interferers are symbol-synchronous to the desired user signal, i.e., $\tau_1 = \tau_2 = 0$. On the other hand, when $\tau_1 = \tau_2 = 0.5T$, the asymptotic maximum MSE reduction ratio attains its smallest possible value. This means that we

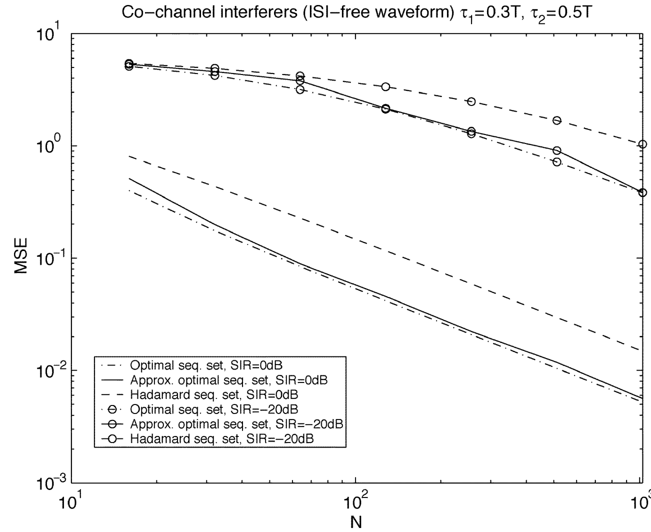


Fig. 4. Comparison of MSEs obtained by using different training sequence sets. Two cochannel interferers with ISI-free waveform waveforms and delays $\tau_1 = 0.3T$, $\tau_2 = 0.5T$.

can almost completely eliminate the effect of the interferers by using the set of long optimal training sequences.

As before, we compare the MSEs of the BLUE channel estimator with the three different training sequence sets in Fig. 4 by considering the case in which $\tau_1 = 0.3T$ and $\tau_2 = 0.5T$. The roll-off factor of the ISI-free waveform is chosen to be $\beta = 0.5$. The other parameters are chosen as in the AR jammer example before. The conclusions from Fig. 4 are similar to those for the rectangular waveform.

V. CONCLUSION

We have solved the problem of optimal training sequence design for MIMO systems over flat fading channels in the presence of colored interfering signals. In order to obtain the advantage of the optimal training sequence design, we have also developed an information feedback scheme that requires a minimal amount of information from the receiver to approximately construct the optimal training sequence set.

Numerical results show that the MSE of the channel estimator at the transmitter can be significantly reduced by using the optimized training sequence set over the Hadamard training sequence set that is often used in the case of white noise. In addition, we observe that comparable estimation performance can be achieved by using the approximate optimal training sequence set obtained by the proposed feedback scheme.

REFERENCES

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

[2] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, pp. 585–595, Nov. 1999.

[3] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1998.

[4] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block coding for wireless communications: performance results," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 452–460, Mar. 1999.

[5] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communication in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 46, pp. 543–564, Mar. 2000.

[6] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1804–1824, July 2002.

[7] S. N. Crozier, D. D. Falconer, and S. A. Mahmoud, "Least sum of squared error (LSSE) channel estimation," *Proc. Inst. Elect. Eng.*, pt. F, vol. 138, pp. 371–378, Aug 1991.

[8] G. Caire and U. Mitra, "Training sequence design for adaptive equalization of multi-user systems," in *Proc. 32nd Asilomar Conf. Signals, Systems and Computers*, vol. 2, Nov. 1998, pp. 1479–1483.

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

[10] C. Tellambura, M. G. Parker, Y. J. Guo, S. J. Shepherd, and S. K. Barton, "Optimal sequences for channel estimation using discrete Fourier transform technique," *IEEE Trans. Commun.*, vol. 47, pp. 230–238, Feb. 1999.

[11] C. Tellambura, Y. J. Guo, and S. K. Barton, "Channel estimation using aperiodic binary sequences," *IEEE Commun. Lett.*, vol. 2, pp. 140–142, May 1998.

[12] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath, "Optimal training for frequency-selective fading channels," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 4, Salt Lake City, UT, May 7–11, 2001, pp. 2105–2108.

[13] A. F. Naguib, V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time coding modem for high data rate wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1459–1478, Oct. 1998.

[14] T. L. Marzetta, "BLAST training: estimating channel characteristics for high capacity space-time wireless," in *Proc. 37th Annual Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Sept. 22–24, 1999.

[15] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inform. Theory*, vol. 49, pp. 951–962, Apr. 2003.

[16] C. Fragouli, N. Al-Dhahir, and W. Turin, "Training-based channel estimation for multiple-antenna broadband transmissions," *IEEE Trans. Wireless Commun.*, vol. 2, pp. 384–391, Mar. 2003.

[17] Y. Song and S. D. Blostein, "Data detection in MIMO systems with co-channel interference," in *Proc. IEEE 56th Vehicular Technology Conf.*, vol. 1, Fall 2002, pp. 3–7.

[18] X. Ma, L. Yang, and G. B. Giannakis, "Optimal training for MIMO frequency-selective fading channels," *IEEE Trans. Wireless Commun.*, to be published.

[19] J. H. Kotecha and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Processing*, vol. 52, no. 2, pp. 546–557, Feb. 2004.

[20] B. Park and T. F. Wong, "Training sequence optimization in MIMO systems with colored noise," in *Proc. IEEE MILCOM*, vol. 1, Boston, MA, Oct. 2003, pp. 135–140.

[21] A. Graham, *Kronecker Products and Matrix Calculus With Applications*. New York: Halsted, 1981.

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

[23] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic, 1968.

[24] T. F. Wong and T. M. Lok, "Transmitter adaptation in multicode DS-CDMA systems," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 69–82, Jan. 2001.

[25] E. K. P. Chong and S. H. Zak, *An Introduction to Optimization*. New York: Wiley, 1996.

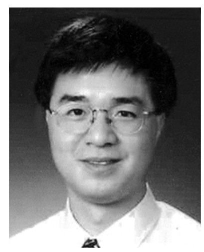
[26] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[27] R. M. Gray. (2002, Aug.) *Toeplitz and Circulant Matrices: A Review* [Online]. Available: http://ee.stanford.edu/~gray/toeplitz.pdf

[28] S. M. Kay, *Modern Spectral Estimation: Theory & Application*. Englewood Cliffs, NJ: Prentice-Hall, 1988.

[29] U. Grenander and G. Szego, *Toeplitz Forms and Their Applications*, 2nd ed. New York: Chelsea, 1984.

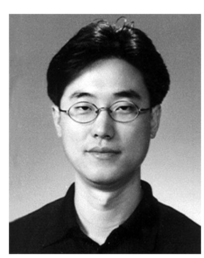
[30] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2001.



Tan F. Wong (S'96–M'98–SM'03) received the B.Sc. degree (with first-class honors) in electronic engineering from the Chinese University of Hong Kong in 1991, and the M.S.E.E. and Ph. D. degrees in electrical engineering from Purdue University, West Lafayette, IN, in 1992 and 1997, respectively.

He was a Research Engineer with the Department of Electronics, Macquarie University, Sydney, Australia, where he was involved with the high-speed wireless networks project. He also served as a Postdoctoral Research Associate with the School of Electrical and Computer Engineering, Purdue University. Since August 1998, he has been with the University of Florida, Gainesville, where he is currently an Associate Professor of electrical and computer engineering.

Prof. Wong serves as the Editor for Wideband and Multiple Access Wireless Systems for the IEEE TRANSACTIONS ON COMMUNICATIONS and as an Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.



Beomjin Park received the B.S. degree in electronic engineering from Hankuk Aviation University, Koyang City, Korea, in 1995, the M.S. degree in electrical engineering from the University of Southern California, Los Angeles, in 1999, and the Ph.D. degree in electrical and computer engineering from the University of Florida, Gainesville, in 2004, in the area of wireless communications.

From 1995 to 1996, he was with Samsung Electronics Company, Ltd., Korea, in the Division of Semiconductors as a Product Engineer. He is currently with Samsung Electronics Company, Ltd., Korea, in the Division of Telecommunication Network as a Research Engineer. His research interests include space–time processing and channel estimation in multiple-input multiple-output systems.