OPTIMAL TRAINING FOR AFFINE-PRECODED AND CYCLIC-PREFIXED BLOCK TRANSMISSIONS

Mounir Ghogho[1] *and Ananthram Swami*[2]
[1] *School of EEE, Univ. of Leeds, LS29JT, United Kingdom*
[2] *Army Research Laboratory, Adelphi, MD 20783-1197, USA*

ABSTRACT

In block-mode transmission, inserting a cyclic prefix (CP) between two consecutive blocks provides an efficient way of removing inter-block interference and simplifies equalization of frequency-selective channels at the receiver. Here, we address optimal training design for channel estimation in such systems. Our investigations focus on affine precoding schemes. Apart from the CP insertion, no redundancy is introduced in the precoding. In designing the precoder, least squares channel estimation is constrained to be decoupled from symbol detection, which results in orthogonal precoding schemes. If the precoding matrix is full-rank, data-rate (or bandwidth) has to be traded off to accommodate training. We propose full-rank orthogonal single-carrier (FROSC) precoding. Then, in order to improve bandwidth efficiency, we propose a *rank deficient* orthogonal single carrier (DROSC) precoding scheme. Symbol recovery is still possible thanks to the finite alphabet property of the data symbols. Finally, simulation results show that the bit error rate of the DROSC method can be quite close to that of the FROSC scheme. The former is however spectrally more efficient.

1. INTRODUCTION

Wireless broadband communications systems are characterized by very dispersive channels. Serial transmissions therefore require complex equalizers. Further, linear zero-forcing equalizers do not exist for such transmissions. To mitigate these problems, block transmission techniques with a guard interval between two consecutive blocks were introduced in [1, 2, 3]. When the duration of the guard interval is at least equal to the delay spread of the channel, interblock interference is avoided and symbol detection can be performed on a per-block basis. Choosing the guard interval to be a cyclic prefix (CP) (i.e. cyclic extension of each block) turns the convolution of the sent data symbols (i.e. outputs of the symbol mapping stage) and the channel into a circular convolution, which simplifies channel equalization [1]. Orthogonal Frequency Division Multiplexing (OFDM) is an example of CP systems, where the data symbols are first linearly precoded by an inverse DFT matrix prior to transmission. If the data symbols are not linearly precoded, the system is commonly referred to as single-carrier CP (SC-CP) system. Comparisons between OFDM and SC-CP have been carried out in many publications e.g. [4, 5, 8] and references therein.

It was shown in [4] that SC-CP with frequency domain linear equalization (FD-LE) outperforms OFDM with fixed modulation, and that OFDM with adaptive modulation (which requires availability of channel state information at the transmitter) outperforms SC-CP with FD-LE. In [5] and [6] among others, it was shown that the use of a decision-feedback equalizer renders SC-CP similar in performance to OFDM with adaptive modulation. The latter however assumes that the channel is quasi static over a large number of blocks as it is often the case for fixed wireless systems such as wireless local area networks. When the channel is fast-varying (e.g. in mobile communications), OFDM with adaptive modulation is not practical. In this case, OFDM must have fixed modulation, i.e. bit/power loading is similar at all subcarriers because the channel is unknown at the transmitter.

When the channel is fast-varying, pilots must be inserted in each block in order to track channel variations. Optimal insertion of pilots for OFDM was investigated in [9] and [10]. Starting from a general affine precoding scheme (see [11] and [12]) and by constraining channel estimation and symbol detection to be decoupled, it was shown in [12] that OFDM with equispaced and equipowered pilot tones is an optimal orthogonal affine precoding scheme in the sense of minimizing the mean square error of channel estimation. In [12], in order to achieve maximum multipath diversity, linearly precoded OFDM (LP-OFDM) was considered. A method which is more efficient in terms of bandwidth utilization was proposed in [13] and it was called linear constellation precoding OFDM (LCP-OFDM). Here, we consider SC-CP whose performance is close to that of maximum diversity and coding gain systems at realistic SNR values. We present explicit design for optimal orthogonal affine-precoded SC-CP.

If the precoding matrix is full-rank, data-rate (or bandwidth) has to be traded off to accommodate training [12]. In the second part of this paper, in order to improve bandwidth efficiency, we propose a *rank deficient* orthogonal single carrier precoding scheme. Symbol recovery is still possible thanks to the finite alphabet property of the data symbols. This method will be shown to be a generalization of the data-dependent superimposed training we have proposed in [17].

Notation: Boldface small (resp. capital) letters denote vectors (resp. matrices.) The $(N \times N)$ DFT matrix is defined as $\mathbf{F} = 1/\sqrt{N}\{\exp(-j2\pi nm/N)\}_{n,m=0}^{N-1}$. The DFT of a vector $\boldsymbol{b}$ is denoted by $\tilde{\boldsymbol{b}} = \mathbf{F}\boldsymbol{b}$. $\mathbf{D}_{\boldsymbol{b}}$ will denote the diagonal matrix whose diagonal is $\boldsymbol{b}$. The nth element of a vector $\boldsymbol{b}$ (or $\mathbf{A}\boldsymbol{b}$ where $\mathbf{A}$ is a matrix)is denoted by $b(n)$ (or $[\mathbf{A}\boldsymbol{b}]_{\mathbf{n}}$). Superscripts *, ^{T} and † denote Hermitian, transpose and pseudo-inverse operators. The Frobenius norm, trace, rank and statistical expectation are denoted by $\|\cdot\|_2$, trace$\{\cdot\}$, rank$\{\cdot\}$ and $E\{\cdot\}$. The $(N \times N)$ identity matrix is denoted by $\mathbf{I}$, and $\mathbf{I}_{N,L}$ will denote the leading $N \times L$ submatrix of $\mathbf{I}_N$. If $\mathcal{P}$ denotes an index set consisting of P elements from $\{0, ..., N-1\}$, then $\mathbf{F}_{\mathcal{P}}$ will denote the $(P \times N)$ submatrix obtained by the $n \in \mathcal{P}$ rows of $\mathbf{F}$.

2. SIGNAL MODEL

Consider a block transmission system where a cyclic prefix (CP), whose duration is longer than the delay spread of the multipath channel, is inserted between two consecutive blocks in order to avoid interblock interference. Assume that the frequency-selective channel is time-invariant over a single block but a frame of K ($K \geq 1$) blocks, where K is less than the channel coherence time in blocks. Let $\boldsymbol{h} = [h_0 \cdots h_{L-1}]^T$ denote the baud-rate sampled discrete-time channel impulse response (CIR) during the frame. After removing the CP, the ith block of the received frame can be modelled as

$$\boldsymbol{x}_i = \mathbf{H}\boldsymbol{u}_i + \boldsymbol{v}_i, \qquad i = 0, 1, ..., K-1, \tag{1}$$

where $\boldsymbol{u}_i$ is the ith transmitted block, and $\mathbf{H}$ is an $(N \times N)$ circulant matrix with first column $[h_0 \ h_1 \ \cdots \ h_{L-1} \ 0 \cdots 0]^T$ and $\boldsymbol{v}_i$ is an AWGN vector with covariance $\sigma^2 \mathbf{I}_N$. Since circular matrices can be diagonalized using DFT matrices, the above signal model can also be rewritten as

$$\boldsymbol{x}_i = \mathbf{F}^* \mathbf{D}_{\tilde{\boldsymbol{h}}} \mathbf{F} \boldsymbol{u}_i + \boldsymbol{v}_i, \quad i = 0, 1, ..., K-1 \tag{2}$$

where $\tilde{\boldsymbol{h}} = [\tilde{h}_0 \cdots \tilde{h}_{N-1}]^T$ with $\tilde{h}_n$ being the frequency response of the channel at the normalized frequency n/N, i.e., $\tilde{h}_n = \sum_{\ell=0}^{L-1} h_\ell \exp(-j2\pi n\ell/N)$.

In order to estimate the frequency-selective channel, training pilots are inserted in each block. We consider non-redundant affine precoding [11]

$$\boldsymbol{u}_i = \boldsymbol{\Theta}_i \boldsymbol{s}_i + \boldsymbol{b}_i \tag{3}$$

where $\boldsymbol{\Theta}_i$ is the $(N \times N)$ precoding matrix in the ith block, $\boldsymbol{s}_i$ is the ith information block and $\boldsymbol{b}_i$ is a known training sequence. The precoding matrix is allowed to vary across the blocks, which gives us some flexibility compared with the fixed schemes in [11] and [12]. Further, the pattern of the set of these matrices could also be cyclicly hopped across frames of K blocks.

The model in eq. (3) encompasses both time-division multiplexed (TDM) and superimposed training. Note that in the absence of training, affine precoding reduces to linear precoding. Apart from the CP insertion, no redundancy is introduced in the precoding scheme. However, our results could be extended to accommodate redundancy, i.e., by letting Θ be tall. We note that here we relax the assumption of a full rank procoding matrix made in [11, 12].

We make the following assumption:

(A1) The non-zero elements of the $\boldsymbol{s}_i$'s are unknown, independent and identically distributed (i.i.d.), and drawn from a finite set of alphabets $\mathcal{A}$.

The training and data powers at the ith block are defined as

$$\sigma_b^2(i) = (N)^{-1} \|\boldsymbol{b}_i\|_2^2$$

$$\sigma_{\Theta s}^2(i) = (N)^{-1} \text{trace}\{\boldsymbol{\Theta}_i \mathbf{R}_i \boldsymbol{\Theta}_i^*\}$$

where $\mathbf{R}_i = E\{\boldsymbol{s}_i \boldsymbol{s}_i^*\}$, which, under (A1), is diagonal. If $s_i(n) = 0$, the nth diagonal element of $\mathbf{R}_i$ is zero.

We also assume without loss of generality that the total transmitted power in each block is unity, i.e., $\sigma_{\Theta s}^2(i) + \sigma_b^2(i) = 1$. Further constraints on the power distributions will be presented later.

Our design criteria will assume a fixed total training power in the frame

$$\sigma_b^2 = \frac{1}{K} \sum_{i=0}^{K-1} \sigma_b^2(i) \ .$$

3. DECOUPLED CHANNEL ESTIMATION AND SYMBOL DETECTION

Although suboptimal, precoding schemes that decouple channel estimation and symbol recovery significantly simplify the receiver. Moreover, it was shown in [16] that the Cramér-Rao bound on channel estimation is minimized when such a precoder is used, assuming that the precoded symbols are Gaussian.

Conditions that guarantee the decoupling of the least squares (LS) channel estimator and symbol detector and optimal training designs were presented in [12]. In this section, we extend those results to the case where $K > 1$ and where the precoding matrix may vary across the blocks. Using eq. (3), the signal model in eq. (1) can be written as

$$\boldsymbol{x}_i = \mathbf{H}\boldsymbol{\Theta}_i \boldsymbol{s}_i + \mathbf{B}_i \boldsymbol{h} + \boldsymbol{v}_i, \quad i = 0, ..., K-1 \tag{4}$$

where $\mathbf{B}_i$ is the leading $(N \times L)$ submatrix of the circular matrix $\mathbf{B}_i^c$ with first column $\boldsymbol{b}_i$, i.e., $\mathbf{B}_i = \mathbf{B}_i^c \mathbf{I}_{N,L}$. Stacking the K received blocks in a single vector $\mathbf{x}_K = [\boldsymbol{x}_0^T \cdots \boldsymbol{x}_{K-1}^T]^T$ we get

$$\mathbf{x}_K = \mathbf{H}_K \boldsymbol{\Theta}_K \mathbf{s}_K + \mathcal{B}_K \boldsymbol{h} + \mathbf{v}_K \tag{5}$$

where $\mathbf{H}_K$ is the $(NK \times NK)$ block diagonal matrix whose diagonal blocks are all equal to $\mathbf{H}$, $\boldsymbol{\Theta}_K$ is defined similar to $\mathbf{H}_K$, $\mathcal{B}_K = [\mathbf{B}_0^T \cdots \mathbf{B}_{K-1}^T]^T$, and $\mathbf{s}_K$ and $\mathbf{v}_K$ are defined similar to $\mathbf{x}_K$. Treating $\mathbf{H}_K \boldsymbol{\Theta}_K \boldsymbol{s}_K$ as an additive noise term, the channel is identifiable if and only if

(C1) $\qquad \text{rank}\{\mathcal{B}_K\} = L$.

Next, we give the frequency domain counterpart of (C1). Let

$$\gamma_n = \sum_{i=0}^{K-1} |\tilde{b}_i(n)|^2, \quad n = 0, ..., N-1$$

where, as defined in Section 1, $\tilde{\boldsymbol{b}}_i$ denotes the DFT of $\boldsymbol{b}_i$. Let P denote the number of nonzero entries of $\boldsymbol{\gamma} := [\gamma_0 \cdots \gamma_{N-1}]$. It is straightforward to show that $\text{rank}\{\mathcal{B}_K\} = \min(P, L)$. Hence, channel identifiability is guaranteed if the combined training power across the blocks is non-zero at at least L frequencies. There is a multitude of training designs that satisfy (C1); for example, for a fixed placement of the P pilot frequencies, the training power can be entirely allocated to a single training block, or it can be (evenly) distributed across the blocks.

When (C1) is satisfied, the LS estimator of $\boldsymbol{h}$ is obtained as

$$\hat{\boldsymbol{h}} = \mathcal{B}_K^\dagger \mathbf{x}_K = \boldsymbol{h} + \mathcal{B}_K^\dagger (\mathbf{H}_K \boldsymbol{\Theta}_K \mathbf{s}_K + \mathbf{v}_K) \ . \tag{6}$$

In order for the LS channel estimator not to be affected by $\mathbf{s}_K$, for any channel realization, the following has to hold (see Appendix 1 for details)

$$\sum_{i=0}^{K-1} \tilde{b}_i^*(n) [\mathbf{F}\boldsymbol{\Theta}_i \boldsymbol{s}_i]_n = 0, \qquad \forall n \ .$$

In order for the above condition to be satisfied for all realizations of the data symbols, the $\boldsymbol{s}_i$'s, that satisfy assumption (A1), the following should be enforced (see Appendix 1 for details)

$$\textbf{(C2)} \qquad \tilde{b}_i^*(n)\ [\mathbf{F}^*\boldsymbol{\Theta}_i \boldsymbol{s}_i]_n = 0, \qquad \forall n, i\ .$$

Condition (C2) implies that if $P \geq L$ then in the i-th block, the DFT of the precoded data $\boldsymbol{\Theta}_i \boldsymbol{s}_i$ should be zero at the activated pilot frequencies, i.e., pilot frequencies carrying power in the ith block. Such precoders will be refereed to as orthogonal precoders as in [12]. The following Lemma gives the optimal training design under (C1) and (C2).

Result 1 *Assume that $Q = N/P$ is an integer. Under the constraint of fixed training power σ_b^2, the mean square error (MSE) of the LS channel estimator in orthogonal precoders is minimized when*

$$\textbf{(C3)} \qquad \gamma_n = \frac{\sigma_b^2 N}{P} \sum_{\ell=0}^{P-1} \delta(n - \ell Q - t)$$

where t is an arbitrary but fixed integer from $[0, ..., Q-1]$

The proof is omitted here because of lack of space. Under (C3), $\mathcal{B}_K^* \mathcal{B}_K = \sigma_b^2 \mathbf{I}_L$, the LS estimate becomes

$$\hat{\boldsymbol{h}} = \frac{1}{\sigma_b^2} \mathcal{B}_K^* \mathbf{x}_K = \frac{1}{\sigma_b^2} \sum_{i=0}^{K-1} \mathbf{B}_i^* \boldsymbol{x}_i$$

and the minimum MSE is found to be

$$\text{mmse}(\hat{\boldsymbol{h}}) := \sum_{\ell=0}^{L-1} E\left\{|\hat{h}_\ell - h_\ell|^2\right\} = \frac{L\sigma^2}{\sigma_b^2}$$

We make the following remarks

- the minimum MSE of $\hat{\boldsymbol{h}}$ is independent of P. Therefore, we will choose $P = L$ in what follows in order to maximize the bandwidth efficiency.
- Result 1 implies that the pilot frequencies should be equispaced and that their *average* powers across the K blocks should be identical. Therefore, channel estimation performance is the same regardless of the distribution of the training power across the blocks.
- The TDM scheme is not an orthogonal precoding scheme. Indeed, condition (C2) implies that training should be superimposed on the data in the time domain (but orthogonal in the frequency domain).
- Result 1 generalizes the findings of [12] where minimization of the channel estimation MSE was carried out for a single block, i.e., $K = 1$. The $K > 1$ scenario gives more flexibility for designing precoders. It also explains why the hopping scheme of [12] works.

4. FULL-RANK ORTHOGONAL PRECODING

In this section, the matrices $\boldsymbol{\Theta}_i$, $i = 1, ..., K-1$, are assumed full-rank. This is the conventional choice for $\boldsymbol{\Theta}$ see e.g., [12, 11]. For example, in the absence of training, the precoding in (3) becomes linear; in this case $\boldsymbol{\Theta}_i = \mathbf{I}$ for single-carrier systems, and $\boldsymbol{\Theta}_i = \mathbf{F}^*$ for OFDM.

4.1. Precoding Design

Let $\mathcal{P}_i$ denote the set of activated pilot frequencies in the ith block, and let P_i denote its cardinality. Since condition (C2) states that in each block, P_i entries of $\mathbf{F}\boldsymbol{\Theta}_i \boldsymbol{s}_i$ should be zero, and since both $\mathbf{F}$ and $\boldsymbol{\Theta}_i$ are full rank, the degrees of freedom of $\boldsymbol{s}_i$ are reduced from N to $N - P_i$. Indeed, $\boldsymbol{s}_i$ should now satisfy P_i linear and independent equations with constant coefficients. Further, the orthogonality equations should be valid for all realizations of the vectors $\boldsymbol{s}_i$. This combined with assumption (A1) implies that for full-rank orthogonal precoding schemes P_i entries of $\boldsymbol{s}_i$ should be set to zero, with corresponding conditions on $\boldsymbol{\Theta}_i$ as discussed below. Without loss of generality, these zeros are chosen to coincide with the pilot index set $\mathcal{P}_i$. The design of the precoding matrix is given next

Result 2 *If $\boldsymbol{\Theta}_i$, $i = 1, ..., K-1$, are full rank and assumption (A1) holds, then the orthogonality condition (C2) is satisfied if and only if the nth entry of $\boldsymbol{E}_i \boldsymbol{s}_i$ is identically zero if $n \in \mathcal{P}_i$, where $\boldsymbol{E}_i$ is any $(N \times N)$ permutation matrix. The latter can be set to the identity matrix without loss of generality. In this case, the precoding matrix has the following form*

$$\boldsymbol{\Theta}_i = \mathbf{F}^* \left(\mathbf{T}_i \mathbf{W}_i \mathbf{T}_i + (\mathbf{I} - \mathbf{T}_i)\mathbf{A_i}\right) \qquad (7)$$

where $\mathbf{T}_i$ an $(N \times N)$ diagonal matrix whose nth diagonal entry is unity if $n \in \mathcal{P}_i$ and zero otherwise, $\mathbf{W}_i$ and $\mathbf{A_i}$ are any $(N \times N)$ matrices such that $(\mathbf{T}_i \mathbf{W}_i \mathbf{T}_i + (\mathbf{I} - \mathbf{T}_i)\mathbf{A_i})$ is full-rank. [1]

By choosing $\mathbf{W}_i = \mathbf{A_i} = \mathbf{I}$, we obtain OFDM with reserved pilot tones, i.e., $\boldsymbol{\Theta}_i = \mathbf{F}^*$. Uncoded OFDM suffers from multipath (or frequency) diversity loss. In fact, only diversity order one is possible over Rayleigh fading channels [7, 14, 13]. To mitigate this problem, dependence among symbols on different subcarriers is introduced through either Galois field (either block or convolutional) channel coding at the bit level (e.g. [15]) or through LP-OFDM or LCP-OFDM [14, 13]. Here, we focus on SC-CP systems. Although such systems do not have full multipath diversity, their performance at realistic SNR values approaches that of maximum diversity systems. Further, maximum diversity at high SNR can be achieved if the constellations are first rotated prior to SC-CP modulation. However, we will not pursue this option in this paper. We focus on SC-CP systems with linear equalizers. Note that nonlinear decoders such as FD-DFE, ML, and sphere decoding could also be used at the expense of increased decoding complexity.

Conventional SC-CP schemes where $\boldsymbol{\Theta}_i = \mathbf{I}$, do not satisfy the orthogonality condition (C2). Next we present a full-rank orthogonal precoding scheme that preserves maximum diversity and coding gains and which can be seen as a modification of SC-CP to satisfy condition (C2). We wish to design an orthogonal precoding matrix $\boldsymbol{\Theta}_i$ such that its nth row is the same as that of the $(N \times N)$ identity matrix if $n \notin \mathcal{P}_i$. This ensures that the information-bearing symbols are not precoded, as in SC-CP systems. We refer to this scheme as full-rank orthogonal single carrier (FROSC) precoding. After some algebra, the precoding matrix is obtained by choosing $\mathbf{W}_i$ and the non-zero $((N - P_i) \times N)$ submatrix of $(\mathbf{I} - \mathbf{T}_i)\mathbf{A_i}$ as

[1] Equivalently, the submatrix of $\mathbf{A_i}$ obtained by deleting the P_i rows and columns corresponding to the active pilot frequencies, should be non-singular, and the submatrix of $\mathbf{W}_i$ correspoding to the pilot frequencies should also be non-singular.

follows (see Appendix 3 for details)

$$\mathbf{W}_i = \mathbf{I}, \quad \text{and} \quad \bar{\mathbf{A}}_i = (\check{\mathbf{F}}^*_{\mathcal{D}_i})^{-1}(\mathbf{I}_{\mathcal{D}_i} - \mathbf{F}^{*T}_{\mathcal{D}_i}\mathbf{T}_i) \qquad (8)$$

where $\mathcal{D}_i$ is the $(N-P_i)$-entry subset of $\{0,...,N-1\}$ indexing the positions of the data frequencies in the ith block ($\mathcal{P}_i \cup \mathcal{D}_i = \{0,...,N-1\}$), $\check{\mathbf{F}}_{\mathcal{D}_i}$ is the $(N-P_i)\times(N-P_i)$ obtained from the nth columns of $\mathbf{F}_{\mathcal{D}_i}$ with $n \in \mathcal{D}_i$, and $\mathbf{I}_{\mathcal{D}_i}$ is the $(N-P_i)\times N$ submatrix obtained from the nth rows of $\mathbf{I}$ with $n \in \mathcal{D}_i$. The resulting precoding matrix is the same as $\mathbf{I}$ except for P_i rows which can be obtained using eqs. (7) and (8).

The bandwidth efficiency in the i-th block is given by

$$\zeta_{OFDM}(i) = \zeta_{FROSC}(i) = \frac{N-P_i}{N+L-1}$$

4.2. Symbol Detection

Once the channel has been estimated using the LS approach in Section III, optimum symbol detection performance is obtained using ML decoding on a per-block basis. Here, in order to reduce decoding complexity, we limit our performance analysis to the MMSE equalizer and symbol-by-symbol detection. Since $\mathbf{H}$ is circulant, equalization can be carried out in the frequency domain. The detected symbols are obtained as

$$\bar{\boldsymbol{s}}_i = \lfloor \mathbf{T}_s\mathbf{F}^*\mathbf{G}\tilde{\boldsymbol{x}}_i \rfloor_{\mathcal{A}} \qquad (9)$$

where $\mathbf{G}$ is the MMSE equalizer which is given by the $(N\times N)$ diagonal matrix whose nth entry $G(n)$ is $G(n) = \hat{\tilde{h}}^*_n/(|\hat{\tilde{h}}_n|^2 + \sigma_v^2)$ with $\hat{\tilde{h}}_n$ being the LS estimate of $\tilde{h}_n$, $\mathbf{T}_s$ is the $((N-P_i)\times N)$ data selection matrix, and $\lfloor \boldsymbol{a} \rfloor_{\mathcal{A}}$ denotes the vector of constellation points from $\mathcal{A}$ that are the closest to the vector $\boldsymbol{a}$.

5. RANK-DEFICIENT ORTHOGONAL PRECODING

Here, we design a precoding scheme with full data-rate. This implies that condition (C2) must be satisfied for all vectors $\boldsymbol{s}_i$ where none of the entries is zero. This is satisfied only when the nth row, $n \in \mathcal{P}_i$, of $[\mathbf{F}\boldsymbol{\Theta}_i]$ is $\mathbf{0}$. Since $\mathbf{F}$ is a full rank matrix, we have that $\text{rank}\{\boldsymbol{\Theta}_i\} = N - P_i$. Rank-deficient precoding matrices prevent linear recovery of the data-symbols. However, using the finite-alphabet property of the latter, detection is still possible as we will see in the next subsection. We design the optimal matrices $\boldsymbol{\Theta}_i$ by minimizing the averaged (over the i.i.d. symbols $\boldsymbol{s}_i$) Euclidean distance between $\boldsymbol{\Theta}_i\boldsymbol{s}_i$ and $\boldsymbol{s}_i$, i.e., we design a rank-deficient orthogonal precoding that is closest to the SC-CP scheme. This leads to the following criterion

$$\min_{\boldsymbol{\Theta}_i;\ \mathbf{F}_{\mathcal{P}_i}\boldsymbol{\Theta}_i = \mathbf{0}} \sum_{i=0}^{K-1} \|\boldsymbol{\Theta}_i - \mathbf{I}\|_2^2 \ .$$

For a fixed $\mathcal{P}_i$, we prove that $\|\boldsymbol{\Theta}_i - \mathbf{I}\|_2$ is minimized when

$$\boldsymbol{\Theta}_i = \mathbf{F}^*(\mathbf{I} - \mathbf{T}_i)\mathbf{F} \ . \qquad (10)$$

The proof is omitted here because of lack of space.

Equation (10) implies that $\boldsymbol{\Theta}_i\boldsymbol{s}_i$ is obtained by first computing the DFT of $\boldsymbol{s}_i$, then the DFT sequence is nulled at the P_i active pilot frequencies, and finally an IDFT is performed.

It is straightforward to show that when $\boldsymbol{\Theta}$ is given by eq. (10), $\|\boldsymbol{\Theta}_i - \mathbf{I}\|_2$ increases with P_i. Therefore, the P_i's should be as small as possible. For example if $L = KM$, then the optimal distribution of the $\mathcal{P}_i$'s is $P_0 = \cdots = P_{K-1} = M$. This is in contrast with the full-rank orthogonal precoding schemes where the distributions of the P_i's (the number of active pilot frequencies) across the blocks was irrelevant. Further, the active pilot frequencies are chosen to be equispaced and (circularly) separated by maximal spacing since this scheme minimizes the MSE of the LS channel estimate.

Combining the above results with the training sequence design, we obtain the following rank-deficient orthogonal single carrier (DROSC) precoding scheme:

Result 3 *Assume $N/L = Q$ and $M = L/K$ are integers. A bandwidth efficient orthogonal precoding scheme is obtained as follows*

- *for $i = 0,...,K-1$ chose $\mathcal{P}_i = \{nKQ + iQ$ for $n = 0,...,M-1$*
- *set the precoding matrices as in eq. (10)*
- *add a training sequence according to condition (C3).*

The above result states that the pilots in each block are equispaced by KQ, and offset by Q from those in the next block. For each block, the precoding consists of first computing the DFT of $\boldsymbol{s}_i$, then the DFT at the P_i active pilot frequencies are set to known pilot values according to Lemma 1, and finally an IDFT is performed, i.e., the elements of $\boldsymbol{\Theta}_i\boldsymbol{s}_i$ are given by

$$[\boldsymbol{\Theta}_i\boldsymbol{s}_i]_n = s_i(n) - e_i(n) \qquad (11)$$

$$e_i(n) = \frac{P_i}{N}\sum_{m=0}^{N-1} s_i(m)\psi_i(n-m) \qquad (12)$$

$$\text{where} \qquad \psi_i(n) = \frac{1}{P_i}\sum_{k\in\mathcal{P}_i} e^{j2\pi nk/N} \ .$$

In Result 3, we assumed $K \leq L$. If $K \geq L$, then at most one entry of the DFT of $\boldsymbol{s}_i$, corresponding to one of the pilot frequencies, should be set to a known value.

DROSC can be seen as data-dependent superimposed training scheme since $\boldsymbol{u}_i = \boldsymbol{s}_i - \boldsymbol{e}_i + \boldsymbol{b}_i$ where $\boldsymbol{e}$ is data-dependent. This method was presented in [17] without any proof of optimality. Here, we have not assumed any a priori structure for $\boldsymbol{\Theta}$ and have shown that the data-dependent superimposed method in [17] is an optimal rank-deficient orthogonal precoding scheme with full multipath diversity. The method in [17] was limited to $K = 1$. When $K > 1$, we show in the simulation section that further improvement in terms of BER can be obtained. If $K = 1$, then $M = L$, and dropping the block index, $\mathcal{P} = \{\ell N/L, \ \ \ell = 0,...,L-1\}$, and $\psi(n) = \frac{L}{N}\delta(n)$. In this case, $e(n)$ is periodic in n with period L, and is given by the cyclic mean of $\boldsymbol{s}_i$, i.e.,

$$e(k+jL) = e(k) = \frac{L}{N}\sum_{m=0}^{Q-1} s(k+mL), \ \ k = 0,...,L-1,$$

which is the scheme presented in [17] and [18].

DROSC is more bandwidth efficient than FROSC since the data rate is not reduced to accommodate training. The bandwidth efficiency of DROSC is given by

$$\zeta_{DROSC} = \frac{N}{N+L-1} \ .$$

The price paid by this increase in bandwidth efficiency is a distortion (or perturbation) of the data symbol sequence by the data dependent sequence $\boldsymbol{e}_i$. However, this effect will fade away when $Q_i := N/P_i$ is large. This combined with the finite alphabet property of the symbols allows for successful symbol recovery as detailed next.

5.1. Symbol Detection

Although nonlinear detectors such as ML and sphere decoding could be used for symbol detection, here, in order to reduce decoding complexity, we only consider the MMSE equalizer and symbol-by-symbol hard detection. The zero-forcing equalizer when applied to SC-CP has poor performance and therefore will not be considered here. Using eq. (10), we can write $\boldsymbol{\Theta}_i \boldsymbol{s}_i = (\mathbf{I} - \mathbf{J}_i)\boldsymbol{s}_i$ where $\mathbf{J}_i = \mathbf{F}^* \mathbf{T}_i \mathbf{F}$. After the channel is estimated, we can remove the contribution of $\boldsymbol{b}_i$ from $\boldsymbol{x}_i$ by simply computing

$$\boldsymbol{z}_i = (\mathbf{I} - \mathbf{J}_i)\boldsymbol{x}_i \; . \tag{13}$$

In the frequency domain, this is equivalent to setting the DFT of $\boldsymbol{x}_i$ at the activated pilot frequencies in the ith block to zero. Since both $\mathbf{H}$ and $\mathbf{J}_i$ are circulant, and hence commutative, $\boldsymbol{z}_i$ can be expressed as

$$\boldsymbol{z}_i = \mathbf{H}(\mathbf{I} - \mathbf{J}_i)\boldsymbol{s} + (\mathbf{I} - \mathbf{J}_i)\boldsymbol{v}_i \; ,$$

where we have used the fact that $(\mathbf{I} - \mathbf{J}_i)^2 = (\mathbf{I} - \mathbf{J}_i)$. The additive noise, $\breve{\boldsymbol{v}}_i = (\mathbf{I} - \mathbf{J}_i)\boldsymbol{v}_i$, is now slightly colored. However, this color will fade away when Q_i is large, and will therefore be ignored in what follows. Further, it is straightforward to show that the power of $\breve{\boldsymbol{v}}_i$ is $\breve{\sigma}_v^2(i) = \sigma_v^2(1 - 1/Q_i)$.

Since $\mathbf{H}$ is circulant, equalization can be carried out in the frequency domain, i.e., the equalized signal is given by

$$\boldsymbol{w}_i = \mathbf{F}^* \mathbf{G} \tilde{\boldsymbol{z}}_i \tag{14}$$

where $\tilde{\boldsymbol{z}}_i$ is obtained by setting $\boldsymbol{z}_i$, the DFT of $\boldsymbol{x}_i$, to zero at the activated pilot frequencies in the ith block. $\mathbf{G}$ is the MMSE equalizer which is given by the $(N \times N)$ diagonal matrix whose nth entry $G(n)$ is $G(n) = \hat{\hat{h}}_n^* / (|\hat{\hat{h}}_n|^2 + \breve{\sigma}_v^2(i))$.

Due to data distortion at the transmission, $\boldsymbol{w}_i \neq \boldsymbol{s}_i$ even in the absence of channel estimation error and noise. Indeed, in this ideal scenario, $\boldsymbol{w}_i = (\mathbf{I} - \mathbf{J}_i)\boldsymbol{s}_i$. Since $(\mathbf{I} - \mathbf{J}_i)$ is singular, $\boldsymbol{s}_i$ cannot be recovered *linearly*. However, since the data symbols are drawn from a finite alphabet and $\mathbf{J}_i \boldsymbol{s}_i$ is small compared to $\boldsymbol{s}_i$, symbol detection can be accomplished by finding the vector of constellation points $\boldsymbol{s}_i$ that minimizes the Euclidian distance between $\boldsymbol{w}_i$ and $(\mathbf{I} - \mathbf{J}_i)\boldsymbol{s}_i$. However, this sequence detection scheme is computationally cumbersome and will therefore not be considered here. Instead, we propose the following iterative symbol-by-symbol detection scheme.

The symbol-by-symbol detection algorithm is initialized by treating $\mathbf{J}_i \boldsymbol{s}_i$ as an extra additive noise term, and considering $\boldsymbol{w}_i$ in eq. (14) as a soft detector of $\boldsymbol{s}_i$; the initial hard detector of $\boldsymbol{s}_i$ is given by

$$\bar{\mathbf{s}}_i^{(0)} = \lfloor \boldsymbol{u}_i \rfloor_{\mathcal{A}}$$

The detected symbols are used to estimate $\mathbf{J}_i \boldsymbol{s}_i$ to be used in the next iteration. The detected symbols at the mth iteration are given by

$$\bar{\mathbf{s}}_i^{(m)} = \lfloor \boldsymbol{w}_i + \mathbf{J}_i \bar{\boldsymbol{s}}_i^{(m-1)} \rfloor_{\mathcal{A}}$$

As we will see in the next section, the symbol detection algorithm converges in only one or two iterations.

6. SIMULATION RESULTS

Here, we investigate the performance of the proposed FROSC and DROSC schemes in terms of bit error rate (BER). The methods are also compared with (uncoded) OFDM. Only decoding based on the MMSE equalizer followed by symbol-by-symbol detection is considered. The length of the data block is set to $N = 64$. The length of the channel is set to $L = 16$. We assume that the channel is time-invariant over $K = 4$ blocks. The channel is randomly generated at each Monte-Carlo run and is assumed to be Rayleigh; the coefficients are uncorrelated and their powers are given by the exponential delay profile $E\left\{|h_\ell|^2\right\} = \exp(-0.2\ell)$. The training sequences are chosen to be optimal according to Lemma 1 and Result 3 and their powers are set to the optimal value given in [12, eq.(24)]. The data symbols are drawn from either BPSK or QPSK constellations. We use the same training power for all precoding schemes; thus, channel estimation performance is identical for all three techniques. We evaluate the BER using 5000 Monte-Carlo runs. Figures 1 and 2, display the BER versus the average energy-per-bit-to-noise ratio (E_b/N_0) for the different schemes in the case of BPSK and QPSK inputs. As expected FROSC outperforms DROSC. However, the latter has a higher bandwidth efficiency than the former. Further, the difference in performance between FROSC and DROSC is negligible if $L/K \ll N$; see Figure 3 where $N = 128$.

7. CONCLUSIONS

Two affine precoding schemes that guarantee decoupling between least squares channel estimation and symbol detection were proposed. In the first scheme, which involves a full-rank precoding matrix, data-rate had to be traded off to accommodate training. In the second scheme, the precoding matrix was designed to be rank deficient in order to preserve maximum data rate. Symbol recovery was still possible thanks to the finite alphabet property of the data symbols. Simulation results show that the bit error rate of the two schemes can be quite close to each other. The extension of this work to multiple antenna systems is an interesting direction for future research.

REFERENCES

[1] J. A. C. Bingham, "Multicarrier modulation for data transmission: an idea whose time has come," *IEEE Communications Magazine*, vol. 28, pp. 5-14, May 1990.

[2] G. B. Giannakis, "Filterbanks for blind channel identification and equalization," *IEEE Signal Processing Lett.*, 4, pp. 184-187, June 1997.

[3] A. Scaglione, G. B. Giannakis and S. Barbarossa, "Redundant filterbank precoders and equalizers-Part I: Unification and optimal design," *IEEE Trans. Signal Processing*, vol. 47, pp. 1988-2006, July 1999.

[4] A. Czylwik, "Comparison between adaptive OFDM and single carrier modulation with frequency domain equalization," in *Proc. IEEE 47th Vehicular Technology Conf.*, vol. 2, pp. 865-869, 1997.

[5] N. Benvenuto and S. Tomasin, "On the comparison between OFDM and single carrier modulation with a DFE using a frequency-domain feedforward filter, *IEEE Trans. Communications*, vol. 50, pp. 947-955, June 2002.

[6] J. Louveaux, L. Vandendorpe and T. Sartenaer, "Cyclic prefixed single carrier and multicarrier transmission: bit rate comparison," *IEEE Communications Lett.*, vol. 7, Apr. 2003.

[7] L. J. Cimini Jr., "Analysis and simulation of a digital mobile channel using orthogonal frequency-division multiplexing," *IEEE Trans. Communications*, vol. 33, pp. 665-675, July 1985.

[8] Z. Wang, X. Ma and G. B. Giannakis, "OFDM or single -carrier block transmissions?," *IEEE Trans. Communications*, vol. 52, pp. 380-394, March 2004.

[9] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," *IEEE Trans. Consumer Electron.*, vol. 44, pp. 1122-1128, Aug. 1998.

[10] S. Adireddy, L. Tong and H. Viswanathan, "Optimal placement of training for frequency-slective block-fading channels," *IEEE Trans. Information Theory*, vol. 48, pp. 2338-2352, Aug 2002.

[11] J. H. Manton, I. Y. Mareels, and Y. Hua, "Affine precoders for reliable communications, " in *Proc. ICASSP*, vol. 5, pp. 2741-2744, Istanbul, Turkey, June 2000.

[12] S. Ohno, and G. B. Giannakis, "Optimal training and redundant precoding for block transmissions with application to wirelesss OFDM," *IEEE Trans. Communications*, vol. 50, pp. 2113-2123, Dec 2002.

[13] Z. Liu, Y. Xin and G. B. Giannakis, "Linear constellation precoding for OFDM with multipath diversity and coding gains," *IEEE Trans. Communications*, vol. 51, March 2003.

[14] Z. Wang and G. B. Giannakis, "Complex-field coding for OFDM over fading wireless channels," *IEEE Trans. Information Theory*, vol. 49, pp. 707-720, Mar. 2003.

[15] B. Le Floch, M. Alard and C. Berrou, "Coded orthogonal frequency division multiplex," *Proc. IEEE*, vol. 83, pp. 982-996, June 1995.

[16] A. Vosoughi and A. Scaglione, "The best training depends upon the receiver structure", ICASSP'04, Montreal, Canada, May 2004.

[17] M. Ghogho, D. C. McLernon, E. Alameda-Hernande, and A. Swami, "Channel estimation and symbol detection for block transmission using data-dependent superimposed training", *IEEE Signal Processing Lett.*, 12(3), 226-229, Mar 2005.

[18] M. Ghogho, D.C. McLernon, E. Alameda-Hernandez and A. Swami, "SISO and MIMO Channel Estimation and Symbol Detection using Data-Dependent Superimposed Training," *IEEE Int. Conference on Acoustics, Speech and Signal Processing* (ICASSP'05) 2005, Philadelphia, March 2005.

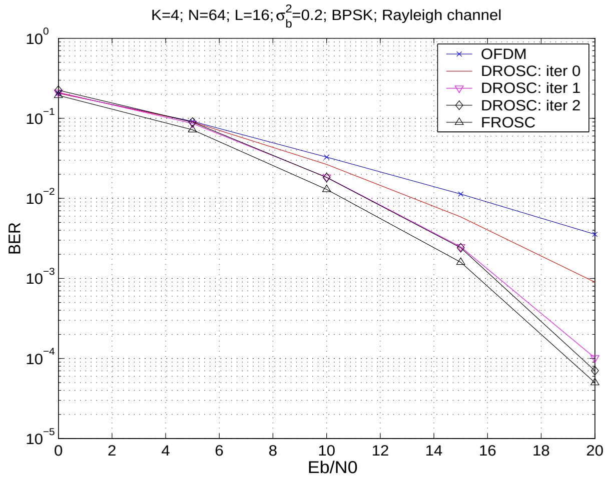

Figure 1. Bit error rate versus E_b/N_0; N=64; BPSK

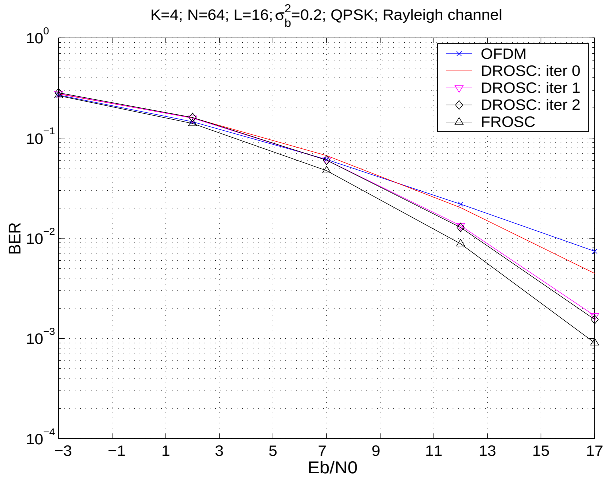

Figure 2. Bit error rate versus E_b/N_0; N=64; QPSK

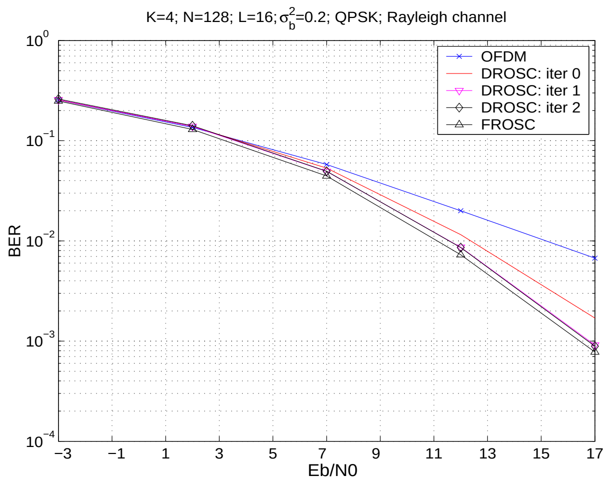

Figure 3. Bit error rate versus E_b/N_0; N=128; QPSK