

# Analysis of A Simplified Channel Estimator for MIMO Frequency Selective Channels

Pei Xiao, Mathini Sellathurai

The Institute of Electronics, Communications and Information Technology

Queen's University Belfast, BT3 9DT, United Kingdom

E-mail: {pei.xiao,m.sellathurai}@ecit.qub.ac.uk

**Abstract**—Channel estimation for multiple-input, multiple-output (MIMO) systems is studied in this paper. In particular, we present a simplified MIMO channel estimator based on orthogonal design. The performance of the proposed scheme is theoretically analyzed and compared to that of the optimum maximum likelihood estimator. The effect of non-orthogonality of the training sequences is investigated. Some modifications of the proposed estimator with sample stacking and averaging are introduced to further improve the estimation performance. This simplified scheme is evaluated in the context of the WiMAX MIMO systems in terms of mean square error for the channel estimation and bit error rate for the space-time turbo equalization. Both analytical and simulation results indicate that despite of its low computational complexity, this simplified estimator leads to minimum variance unbiased estimation and achieves identical performance to that of the maximum likelihood estimator.

**Keywords:** MIMO, space-time codes, channel estimation, frequency-selective fading channels.

## I. INTRODUCTION

In recent years, space-time coding has emerged as one of the most promising technologies for meeting the high data rate and high service quality requirements. Space-time codes were first introduced in [1] to provide transmit diversity in wireless fading channels using multiple antennas. There are two main types of space-time codes, namely, space-time trellis codes (STTC) [1] and space-time block codes (STBC) [2], [3]. In this work, we consider the use of STBC, in particular, the two-antenna transmit diversity scheme [2] proposed by Alamouti.

At high data rates, the intersymbol interference (ISI) introduced by frequency-selective channels becomes a severe problem. An effective remedy is the use of an equalizer, or more effectively, a turbo equalizer that performs equalization and channel decoding jointly in an iterative manner. The performance of a turbo equalizer depends largely on the accuracy of channel estimation. Perfect knowledge of the underlying channel is usually assumed in the design of such systems. However, in a practical system, the channel has to be estimated. The estimation of MIMO channels is more challenging compared to that of SISO channels since a large number of channel parameters have to be estimated. The issue of estimating frequency-selective MIMO channels has been addressed in several papers. For example, Kalman filters are used to track the channel variations in [4], and space-time decoding and channel estimation are performed jointly in spatially correlated time-varying channels. In [5], the redundancy induced by space-time block codes is utilized to blindly identify MIMO channels using the subspace method. The algorithm can be further extended to a semi-blind approach by incorporating pilot symbols. A pilot embedding scheme

is proposed in [6] in order to obtain an initial estimate of the MIMO channels and improved estimates are obtained by integrating the channel estimator into the iterative decoding loop. The design of optimal training sequences for multiple-antenna systems in a dispersive environment was investigated in [7], in which some sequences with good autocorrelation and cross correlation properties were found by exhaustive search and tabulated.

In this paper, we show that the problem of channel estimation can be simplified by the orthogonal design of training sequences, similar to the idea of using the orthogonal structure of the space-time block codes to achieve maximum-likelihood decoding with simple linear processing at the receiver. Our analysis reveals that despite of its lower computational complexity, the simplified scheme does not incur any performance loss compared to the optimum maximum likelihood (ML) estimation. They both attain Cramer-Rao Lower Bound (CRLB), which is the best performance that can be achieved by an unbiased estimator. The channel estimates obtained are provided to a space-time turbo equalizer in order to demonstrate the effectiveness of the channel estimator in terms of the BER performance of the equalizer.

## II. SYSTEM MODEL

To simplify the algorithm derivation, we use a 3-tap WiMAX MIMO channel specified in IEEE 802.16 standard [8] as an example. The algorithm can be easily extended to address generic ISI channels. Tailored for different terrain conditions, a set of 6 typical channel models, known as the Stanford University Interim (SUI) Channel Models have been proposed in [9] for simulation, design, development and testing of WiMAX systems. For the purpose of this study, we select the SUI-3 channel which has 3 taps with a tap spacing of 500ns, and maximum tap delay of 1000ns. Each complex channel coefficient is denoted as  $h_{ij}^l$  where the first (second) subscript  $i(j)$  is the index of the transmit (receive) antenna, the superscript  $l$  refers to the number of the channel tap. For example,  $h_{01}^0$  denotes the channel coefficient corresponding to the first tap of the channel between the transmit antenna  $tx0$  and the receive antenna  $rx1$ . The channel coefficients are assumed to remain constant during the transmission of one block of data. The received signals at antenna  $rx0$  and  $rx1$  can be formed as

$$\begin{aligned} r_n &= h_{00}^2 s_{n-2}^0 + h_{00}^1 s_{n-1}^0 + h_{00}^0 s_n^0 + h_{10}^2 s_{n-2}^1 \\ &\quad + h_{10}^1 s_{n-1}^1 + h_{10}^0 s_n^1 + w_n; \\ y_n &= h_{01}^2 s_{n-2}^0 + h_{01}^1 s_{n-1}^0 + h_{01}^0 s_n^0 + h_{11}^2 s_{n-2}^1 \\ &\quad + h_{11}^1 s_{n-1}^1 + h_{11}^0 s_n^1 + v_n, \end{aligned} \quad (1)$$

where  $w_n, v_n$  are the complex additive white Gaussian noise with zero mean and variance  $N_0$ , and  $s_{n-i}^j$  denotes the QPSK symbol transmitted from the  $j^{\text{th}}$  antenna at time instant  $n-i$ . Based on (1), the received signal can be expressed in matrix form as  $\mathbf{R}_n = \mathbf{S}_n \mathbf{H} + \mathbf{V}_n$ , where

$$\begin{aligned} \mathbf{R}_n &= \begin{bmatrix} r_n & y_n \\ r_{n+1} & y_{n+1} \end{bmatrix}, \quad \mathbf{V}_n = \begin{bmatrix} w_n & v_n \\ w_{n+1} & v_{n+1} \end{bmatrix}; \\ \mathbf{S}_n &= \begin{bmatrix} s_n^0 & s_{n-1}^0 & s_{n-2}^0 & s_n^1 & s_{n-1}^1 & s_{n-2}^1 \\ s_{n+1}^0 & s_n^0 & s_{n-1}^0 & s_{n+1}^1 & s_n^1 & s_{n-1}^1 \end{bmatrix}; \\ \mathbf{H} &= \begin{bmatrix} h_{00}^0 & h_{00}^1 & h_{00}^2 & h_{10}^0 & h_{10}^1 & h_{10}^2 \\ h_{01}^0 & h_{01}^1 & h_{01}^2 & h_{11}^0 & h_{11}^1 & h_{11}^2 \end{bmatrix}^T. \end{aligned} \quad (2)$$

### III. CHANNEL ESTIMATION

The task of a channel estimator is to estimate the fading coefficients  $h_{ij}^l$  given the received observation  $\{r_n, y_n\}$ , and some knowledge of the transmitted data. In the derivation of the channel estimation algorithm, we assume that  $h_{ij}^l$  are unknown deterministic and no prior knowledge of  $h_{ij}^l$  is available at the receiver. In this section, we first present the conventional maximum likelihood (ML) channel estimator, which is optimum in the sense that it leads to minimum variance unbiased (MVU) estimation, and attains the Cramer-Rao Lower Bound (CRLB). A simplified channel estimation algorithm with orthogonal training sequences is then presented and is shown to achieve the same performance as the ML estimator.

#### A. Maximum likelihood channel estimator

Let us denote

$$\begin{aligned} \mathbf{r}_n &= [r_n \quad r_{n+1}]^T, \quad \mathbf{y}_n = [y_n \quad y_{n+1}]^T; \\ \mathbf{w}_n &= [w_n \quad w_{n+1}]^T, \quad \mathbf{v}_n = [v_n \quad v_{n+1}]^T; \\ \mathbf{h}_r &= [h_{00}^0 \quad h_{00}^1 \quad h_{00}^2 \quad h_{10}^0 \quad h_{10}^1 \quad h_{10}^2]^T; \\ \mathbf{h}_y &= [h_{01}^0 \quad h_{01}^1 \quad h_{01}^2 \quad h_{11}^0 \quad h_{11}^1 \quad h_{11}^2]^T, \end{aligned} \quad (3)$$

then (2) can be split into two linear vector models  $\mathbf{r}_n = \mathbf{S}_n \mathbf{h}_r + \mathbf{w}_n$ , and  $\mathbf{y}_n = \mathbf{S}_n \mathbf{h}_y + \mathbf{v}_n$ . Given the data matrix  $\mathbf{S}_n$ , the maximum likelihood estimate of the vector  $\mathbf{h}_r$  and  $\mathbf{h}_y$  are [10]

$$\begin{aligned} \hat{\mathbf{h}}_r &= (\mathbf{S}_n^* \mathbf{S}_n)^{-1} \mathbf{S}_n^* \mathbf{r}_n = \mathbf{S}_n^\dagger \mathbf{r}_n; \\ \hat{\mathbf{h}}_y &= (\mathbf{S}_n^* \mathbf{S}_n)^{-1} \mathbf{S}_n^* \mathbf{y}_n = \mathbf{S}_n^\dagger \mathbf{y}_n, \end{aligned} \quad (4)$$

where  $\mathbf{S}_n^\dagger = (\mathbf{S}_n^* \mathbf{S}_n)^{-1} \mathbf{S}_n^*$  denotes the left pseudoinverse of  $\mathbf{S}_n$ . The ML estimator expressed by (4) suffers from a dimensionality problem. When the columns of  $\mathbf{S}_n$  are not statistically independent, the matrix  $\mathbf{S}_n$  will not have full column rank and the described procedure will fail since  $\mathbf{S}_n^* \mathbf{S}_n$  is not invertible. The problem can be tackled by stacking the received vectors  $\mathbf{r}_n, \mathbf{y}_n$  and expanding the matrices, i.e.,

$$\begin{aligned} \mathbf{R} &= \begin{bmatrix} r_n & y_n \\ r_{n+1} & y_{n+1} \\ \vdots & \vdots \\ r_{n+D-1} & y_{n+D-1} \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} w_n & v_n \\ w_{n+1} & v_{n+1} \\ \vdots & \vdots \\ w_{n+D-1} & v_{n+D-1} \end{bmatrix}; \\ \mathbf{S} &= \begin{bmatrix} s_n^0 & s_{n-1}^0 & s_{n-2}^0 & s_n^1 & s_{n-1}^1 & s_{n-2}^1 \\ s_{n+1}^0 & s_n^0 & s_{n-1}^0 & s_{n+1}^1 & s_n^1 & s_{n-1}^1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ s_{n+D-1}^0 & s_{n+D-2}^0 & s_{n+D-3}^0 & s_{n+D-1}^1 & s_{n+D-2}^1 & s_{n+D-3}^1 \end{bmatrix}; \end{aligned} \quad (5)$$

Now, the received signal can be expressed in matrix form as  $\mathbf{R} = \mathbf{S} \mathbf{H} + \mathbf{V}$ . Note that the above signal model is formed by stacking  $D$  received samples from each receive antenna. Denoting

$$\begin{aligned} \mathbf{R} &= [\mathbf{r} \quad \mathbf{y}], \quad \mathbf{V} = [\mathbf{w} \quad \mathbf{v}]; \\ \mathbf{r} &= [r_n \quad \dots \quad r_{n+D-1}]^T, \quad \mathbf{y} = [y_n \quad \dots \quad y_{n+D-1}]^T; \\ \mathbf{w} &= [w_n \quad \dots \quad w_{n+D-1}]^T, \quad \mathbf{v} = [v_n \quad \dots \quad v_{n+D-1}]^T; \\ \mathbf{r} &= \mathbf{S} \mathbf{h}_r + \mathbf{w}, \quad \mathbf{y} = \mathbf{S} \mathbf{h}_y + \mathbf{v}, \end{aligned} \quad (6)$$

the ML estimate of the vector  $\mathbf{h}_r$  and  $\mathbf{h}_y$  are

$$\hat{\mathbf{h}}_r = (\mathbf{S}^* \mathbf{S})^{-1} \mathbf{S}^* \mathbf{r} = \mathbf{S}^\dagger \mathbf{r}; \quad \hat{\mathbf{h}}_y = (\mathbf{S}^* \mathbf{S})^{-1} \mathbf{S}^* \mathbf{y} = \mathbf{S}^\dagger \mathbf{y},$$

where  $\mathbf{S}^\dagger = (\mathbf{S}^* \mathbf{S})^{-1} \mathbf{S}^*$  denotes the left pseudoinverse of  $\mathbf{S}$ . Define  $\hat{\mathbf{H}} = [\hat{\mathbf{h}}_r \quad \hat{\mathbf{h}}_y]$ , we have

$$\begin{aligned} \hat{\mathbf{H}}^{\text{ML}} &= [\mathbf{S}^\dagger \mathbf{r} \quad \mathbf{S}^\dagger \mathbf{y}] = \mathbf{S}^\dagger [\mathbf{r} \quad \mathbf{y}] = \mathbf{S}^\dagger \mathbf{R} \\ &= \mathbf{S}^\dagger (\mathbf{S} \mathbf{H} + \mathbf{V}) = \mathbf{H} + \mathbf{S}^\dagger \mathbf{V}, \end{aligned} \quad (7)$$

which is an unbiased estimate of  $\mathbf{H}$  since  $\mathbb{E}[\hat{\mathbf{H}}] = \mathbf{H} + \mathbf{S}^\dagger \mathbb{E}[\mathbf{V}] = \mathbf{H}$ . In Appendix, it is proven that this ML estimator gives minimum variance unbiased estimation and attains the CRLB, which is the lower bound on the variance of any unbiased estimator. In Appendix, the CRLB is derived as

$$\mathbf{C}_{\hat{\mathbf{h}}} = N_0 \mathbb{E}[(\mathbf{S}^* \mathbf{S})^{-1}] \geq \frac{N_0 \mathbf{I}_{2L}}{2E_b D}. \quad (8)$$

where  $L$  is the number of taps per channel. For the SUI-3 channel,  $L = 3$ . Equality in (8) holds when orthogonal training sequences are used.

Depending on the form of the data that can be retrieved, channel estimation can be either decision-directed or pilot-assisted. The former uses decision feedback loops and utilizes the decisions on the transmitted signals  $\hat{\mathbf{S}}$  to extract the channel coefficients. The second approach makes the use of training sequence (pilot symbols), i.e.,  $\mathbf{S}$  is known in this case. For quasi-static and slowly-varying fading channels, pilot-assisted channel estimation is often used in practice, and it is the approach we take here in order to simplify the estimation process.

#### B. A simplified channel estimator

Recall that the received signal can be expressed in vector form as  $\mathbf{r} = \mathbf{S} \mathbf{h}_r + \mathbf{w}$ . In the case of orthogonal training sequences, the columns of  $\mathbf{S}$  are orthogonal to each other so that  $\mathbf{S}^* \mathbf{S}$  is a diagonal matrix with all the diagonal entries equal to a constant  $\alpha$ , i.e.,  $\mathbf{S}^* \mathbf{S} = \alpha \mathbf{I}$ . According to (5),  $\alpha = E_s D = 2E_b D$ , where  $E_s$  and  $E_b$  refer to symbol and bit energy, respectively. Processing the received vector  $\mathbf{r}$  with  $\mathbf{S}^*$ , yield

$$\mathbf{S}^* \mathbf{r} = \mathbf{S}^* \mathbf{S} \mathbf{h}_r + \mathbf{S}^* \mathbf{w} = 2E_b D \mathbf{h}_r + \mathbf{S}^* \mathbf{w} \quad (9)$$

Utilizing the orthogonality of the matrix  $\mathbf{S}$ , the channel estimate can be simply derived as

$$\begin{aligned} \hat{\mathbf{h}}_r &= \frac{\mathbf{S}^* \mathbf{r}}{2E_b D} = \mathbf{h}_r + \frac{\mathbf{S}^* \mathbf{w}}{2E_b D} \\ \hat{\mathbf{H}} &= [\hat{\mathbf{h}}_r \quad \hat{\mathbf{h}}_y] = \frac{\mathbf{S}^* [\mathbf{r} \quad \mathbf{y}]}{2E_b D} = \frac{\mathbf{S}^* \mathbf{R}}{2E_b D} \end{aligned} \quad (10)$$

which is an unbiased estimate since  $E[\hat{\mathbf{H}}] = \mathbf{H} + \mathbf{S}^* E[\mathbf{V}]/(2E_b D) = \mathbf{H}$ . By comparing (10) with (7), we can see that the proposed scheme does not involve the matrix inverse  $(\mathbf{S}^* \mathbf{S})^{-1}$ . Therefore, it has a lower complexity and does not have dimensionality problem. Next, we will prove that this estimator achieves the same estimation performance as the ML estimator. Assume  $\mathbf{h}_r$  and  $\mathbf{h}_y$  have the same channel statistics, and let  $\mathbf{h}$  represents either  $\mathbf{h}_r$  or  $\mathbf{h}_y$ . The estimation error vector is computed as  $\mathbf{e}_h = \hat{\mathbf{h}} - \mathbf{h}$ . According to (10),  $\mathbf{e}_h = \frac{\mathbf{S}^* \mathbf{w}}{2E_b D}$ , its covariance matrix can be derived as

$$\mathbf{C}_{\hat{\mathbf{h}}} = E[\mathbf{e}_h \mathbf{e}_h^*] = \frac{\mathbf{S}^* E[\mathbf{w} \mathbf{w}^*] \mathbf{S}}{(2E_b D)^2} = \frac{N_0 \mathbf{I}_{2L}}{2E_b D} \quad (11)$$

which is the same as (8) for the ML estimator, meaning that this simple channel estimation scheme also attains CRLB and leads to minimum variance unbiased estimation.

Now, we shall analyze the effect of non-orthogonality on the performance of the proposed estimator. In the case when the training sequences are not orthogonal,  $\mathbf{S}^* \mathbf{S}$  can be decomposed into an identity matrix  $\mathbf{I}$  and a non-identity matrix  $\mathbf{K}$ , i.e.,  $\mathbf{S}^* \mathbf{S} = 2E_b D(\mathbf{I} + \mathbf{K})$ . For example, if an m-sequence is used, the diagonal elements of  $\mathbf{K}$  are all zeros and non-diagonal elements of  $\mathbf{K}$  are all ones. Equation (9) and (10) should be reformed as

$$\begin{aligned} \mathbf{S}^* \mathbf{r} &= \mathbf{S}^* \mathbf{S} \mathbf{h} + \mathbf{S}^* \mathbf{w} = 2E_b D(\mathbf{I} + \mathbf{K}) \mathbf{h} + \mathbf{S}^* \mathbf{w} \\ \hat{\mathbf{h}} &= \frac{\mathbf{S}^* \mathbf{r}}{2E_b D} = \mathbf{h} + \mathbf{K} \mathbf{h} + \frac{\mathbf{S}^* \mathbf{w}}{2E_b D}, \end{aligned} \quad (12)$$

which is no longer an unbiased estimate of  $\mathbf{h}$  since  $E[\hat{\mathbf{h}}] = \mathbf{h} + \mathbf{K} \mathbf{h} + \mathbf{S}^* E[\mathbf{w}]/(2E_b D) = \mathbf{h} + \mathbf{K} \mathbf{h} \neq \mathbf{h}$ . The biased term  $\mathbf{K} \mathbf{h}$  is due to the contribution from channel vector  $\mathbf{h}$  itself (self-interference). The estimation error vector and the mean square error matrix become

$$\begin{aligned} \mathbf{e}_h &= \hat{\mathbf{h}} - \mathbf{h} = \mathbf{K} \mathbf{h} + \frac{\mathbf{S}^* \mathbf{w}}{2E_b D} \\ \mathbf{C}_{\hat{\mathbf{h}}} &= E[\mathbf{e}_h \mathbf{e}_h^*] = \mathbf{K} E[\mathbf{h} \mathbf{h}^*] \mathbf{K}^* + \frac{N_0 \mathbf{I}_{2L}}{2E_b D} \end{aligned} \quad (13)$$

From (12) and (13), it is obvious that the performance of the proposed algorithm is not only affected by the noise but also by the self-interference due to the non-orthogonality of training sequences. As indicated by the numerical results presented in Section IV, the orthogonality of the training sequences is essential for the proposed scheme to achieve good performance. For flat fading channels, Hadamard codes can be applied to satisfy this condition. However, for multipath channels, the training sequence in each transmit antenna not only has to be orthogonal to its shifts within  $V$  taps but also has to be orthogonal to the training sequences in other antennas and their shifts within  $V$  taps, where  $V$  is the order of channel memory [7]. Some optimal PVM codes<sup>1</sup> meeting the above requirements are given in [7] for code length less than or equal to 16. Another alternative is to use the maximum-length shift-register sequences, or m-sequences for short. Each m-sequence is periodic with period  $n = 2^m - 1$ , where  $m$  is a positive integer. The m-sequence is a binary periodic sequence exhibiting a periodic autocorrelation

<sup>1</sup>Denote  $L_t$  as the length of each training sequence,  $M$  as the number of transmit antenna, and  $P = L_t - V$ . A sequence set is called a  $(P, V, M)$  code if the training sequence in each antenna is not only orthogonal to its shift within  $V$  taps but also orthogonal to the training sequences in other antennas and their shifts within  $V$  taps [7].

$\phi(j)$  with values  $\phi(j) = n$  for  $j = 0$ , and  $\phi(j) = -1$  for  $1 \leq j \leq n - 1$  [11]. This impulse-like autocorrelation implies that the power spectrum is nearly white, hence the sequence resembles white noise. Different training sequences including PVM sequences, m-sequences and random sequences are applied to the ML and the proposed channel estimator and their performance compared in Section IV of this paper.

According to (11), the estimation variance for the complex channel gain  $h_{ij}^l$  is  $\text{var}(\hat{h}_{ij}^l) = \frac{N_0}{2E_b D}$ . This means that the estimation error decreases linearly as the stacking factor  $D$  increases. Another way of further improving the estimation performance is to average several independent estimates. Suppose that we have  $N$  independent estimates of  $h_{ij}^l$ , denoted as  $\hat{h}_{ij}^{l1}, \hat{h}_{ij}^{l2}, \dots, \hat{h}_{ij}^{lN}$ . A noise resistant estimate of  $h_{ij}^l$  can be obtained as  $\hat{h}_{ij}^l = \sum_{n=1}^N \hat{h}_{ij}^{ln} / N$ . Assume each of the estimates is derived based on the algorithm describe above. For each  $\hat{h}_{ij}^l$ ,

$$E[h_{ij}^l - \hat{h}_{ij}^{ln}] = 0, \quad E[(h_{ij}^l - \hat{h}_{ij}^{ln})^2] = \frac{N_0}{2E_b D}.$$

Using the fact that the estimates are independent from each other, it can be easily shown that

$$E[h_{ij}^l - \hat{h}_{ij}^l] = E\left[h_{ij}^l - \sum_{n=1}^N \frac{\hat{h}_{ij}^{ln}}{N}\right] = 0$$

$$\begin{aligned} E[(h_{ij}^l - \hat{h}_{ij}^l)^2] &= E\left[\left(h_{ij}^l - \sum_{n=1}^N \frac{\hat{h}_{ij}^{ln}}{N}\right)^2\right] \\ &= \frac{\sum_{n=1}^N E[(h_{ij}^l - \hat{h}_{ij}^{ln})^2]}{N^2} = \frac{N N_0}{2E_b D N^2} = \frac{N_0}{2E_b D N}. \end{aligned} \quad (14)$$

Let us denote  $\bar{\mathbf{C}}_{\hat{\mathbf{h}}}$  as the the sample averaged version of the covariance matrix  $\mathbf{C}_{\hat{\mathbf{h}}}$ . Following the same analysis as described above,  $\bar{\mathbf{C}}_{\hat{\mathbf{h}}}$  can be derived as

$$\bar{\mathbf{C}}_{\hat{\mathbf{h}}} = \frac{N_0 E[(\mathbf{S}^* \mathbf{S})^{-1}]}{N} = \frac{N_0 \mathbf{I}_{2L}}{2E_b D N}, \quad (15)$$

One can see from (15) that averaging several independent channel estimates also has the effect of reducing estimation errors.

#### IV. ANALYTICAL AND SIMULATION RESULTS

The channel estimation algorithms introduced in Section III are evaluated over WiMAX SUI-3 MIMO channels and results are presented in this section to assess their performance. The WiMAX channel coefficients vary from one block to another, however, they are assumed to remain constant during the transmission of one block of data. It is therefore a quasi-static channel. The simulation curves are obtained by averaging the simulation results over 1000 channel realizations. The antenna correlation coefficient is set to 0.4. The noise variance  $N_0$  and path delays are assumed to be known to the receiver.

Fig. 1 shows the comparison between the analytical and simulation results for the ML estimator using random sequence and m-sequence. The employed m-sequence has a degree of  $m = 5$ , and is obtained with the generator polynomial  $g(p) = p^5 + p^3 + 1$ . It is a periodic binary sequence with period  $n = 2^5 - 1 = 31$ . The binary sequence is then converted to a symbol sequence in such a way that 0 is mapped to symbol  $\frac{1}{\sqrt{2}} + j \frac{1}{\sqrt{2}}$ , and 1 is mapped to symbol  $\frac{-1}{\sqrt{2}} - j \frac{1}{\sqrt{2}}$ . The symbol

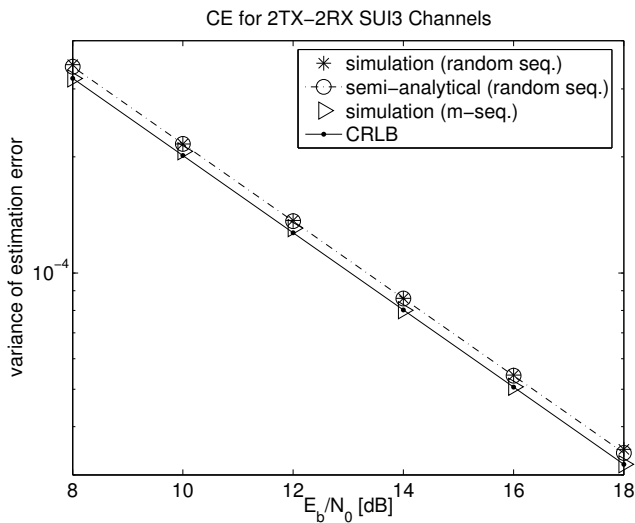


Fig. 1. Performance of the ML channel estimator for the SUI-3 MIMO channels ( $D \times N = 31 \times 12$ ).

sequence and its cyclic shifted version are transmitted from antenna  $tx_0$  and  $tx_1$ , respectively. Note that there has to be enough shift to guarantee low cross correlation between the two antennas. We use the variance of the estimation error as the performance measurement. As shown in Fig. 1, the simulation results for the ML estimator is in close agreement with the CRLB expressed by (15) for the m-sequence. In the case when random QPSK symbols are transmitted as training sequence, the theoretical curve is derived semi-analytically by averaging the matrix  $(\mathbf{S}^* \mathbf{S})^{-1}$  in the simulations, and scaling it with factor  $\frac{N_0}{N}$ . As depicted in Fig. 1, the semi-analytical results coincide with the simulation results. They both exhibit a small discrepancy compared with the CRLB because the matrix  $\mathbf{S}$  in (15) is not strictly orthogonal for random sequences.

Different training sequences are compared in Fig. 2 and Fig. 3 for the ML estimator and the proposed channel estimator, respectively. The chosen PVM sequences are  $X_1 = \{1111010001110\}$  and  $X_2 = \{111101101000\}$ , which are taken from [7]. In this case, the parameter settings are  $P = 12, V = 2, M = 2$ . The binary sequence to symbol sequence conversion process is the same as stated above. It can be seen from Fig. 2 that for the ML channel estimator, the m-sequence yields almost identical performance to the optimal PVM sequence. They both achieve CRLB. The performance of random sequence is slightly worse. However, the gap is very small, the performance loss by using random sequence is only about 0.2 dB. Apparently, the orthogonality is not a major concern for the ML estimator. Fig. 3 shows that with orthogonal PVM sequences, the proposed channel estimator achieves CRLB. Therefore, it does not incur performance loss compared to the ML estimator. However, it performs very poorly when the training sequences are not orthogonal. The curves corresponding to the random sequence and m-sequence are virtually flat. The reason is that with non-orthogonal sequences, the impact of self-interference on the proposed scheme is dominant. In the mean square error matrix expressed by (13), the self-interference term  $\mathbf{K} \mathbf{E}[\mathbf{h}\mathbf{h}^*] \mathbf{K}^*$  is much bigger than the noise term  $\frac{N_0 \mathbf{I}_{2L}}{2E_b D}$ . Unlike the ML scheme, the proposed scheme is very sensitive to the orthogonal condition, it even fails to work with the near-orthogonal m-sequence.

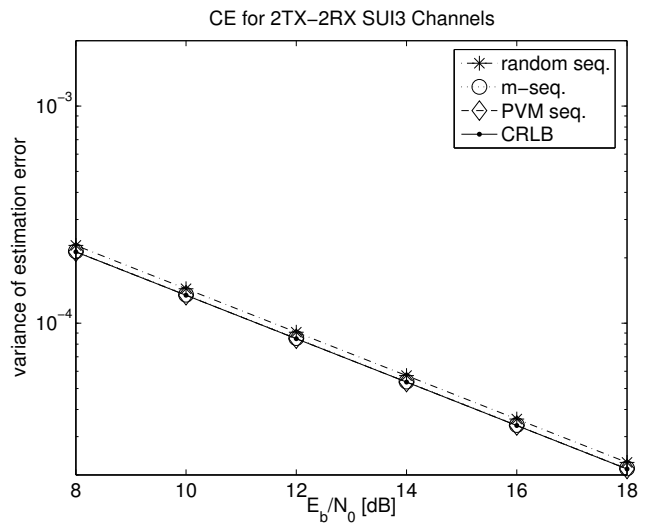


Fig. 2. Comparison of different training sequences for ML channel estimator ( $D \times N = 31 \times 12$ ).

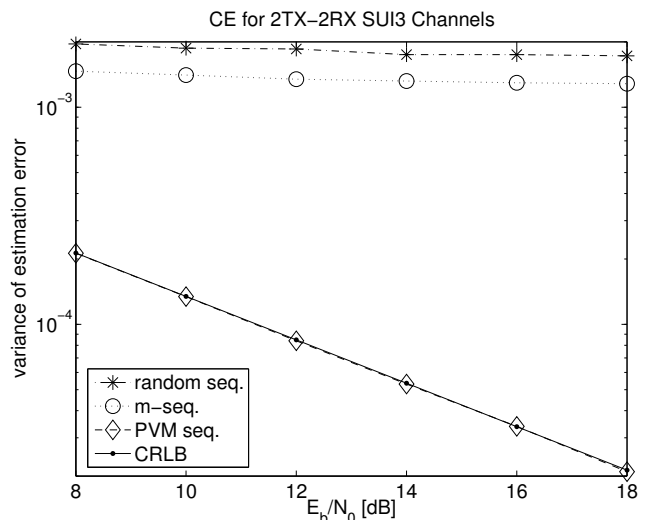


Fig. 3. Comparison of different training sequences for the proposed channel estimator ( $D \times N = 31 \times 12$ ).

The important conclusion is that orthogonality of the training sequence has to be guaranteed in order for the proposed scheme to achieve the desired results.

The effect of the averaging factor is examined for the proposed channel estimator in Fig. 4 using orthogonal PVM sequences. As both analytical and simulation results indicate, the estimation variance decreases as the averaging factor  $N$  increases. A similar behavior is also observed for the stacking factor  $D$  (the plot is omitted here to conserve space). This implies two ways of improving channel estimation performance. For example, at low  $E_b/N_0$ , we can achieve the same performance obtained with a higher  $E_b/N_0$  by increasing the values of  $N$  and/or  $D$ . In addition to noise removal, the stacking method also prevents the dimensionality problem for the ML estimator as mentioned previously. However, the performance improvement comes at the penalty of wasting system resources since increasing the values of  $N$  and/or  $D$  means transmitting more pilot symbols. It also should be noted that the sample averaging technique is only useful for static and slowly time-

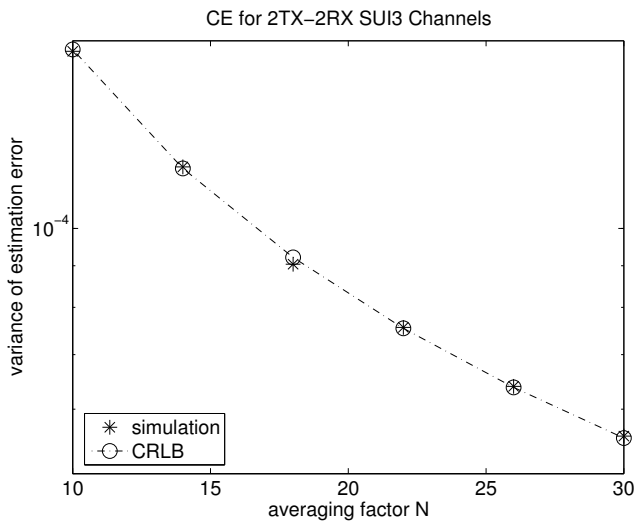


Fig. 4. Impact of the averaging factor  $N$  on the estimation performance ( $E_b/N_0 = 14$  dB,  $D = 12$ ).

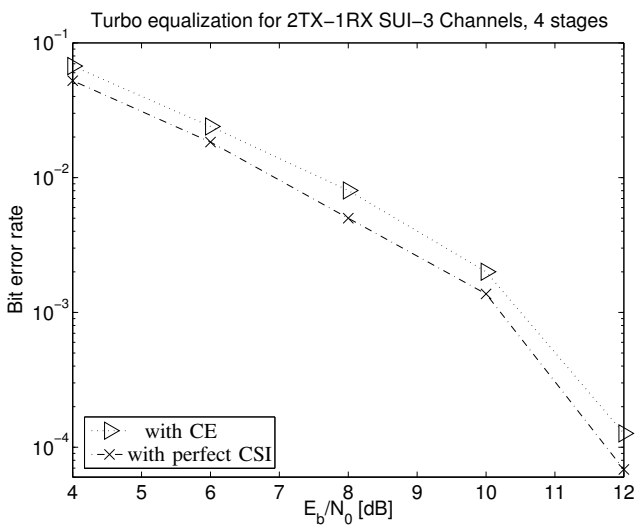


Fig. 5. Performance of the turbo equalization with channel estimate (CE) and perfect CSI for the 2TX-1RX system. The curves represent the 4th stage turbo equalization ( $D \times N = 12 \times 8$ ).

varying (e.g., quasi-static) ISI channels. For fast time-varying channels, some other techniques, e.g., sample smoothing [12], or the Kalman filter [13] can be used.

The performance of the joint channel estimation and turbo equalization is examined in Fig. 5 and Fig. 6 for the 2TX-1RX and 2TX-2RX system, respectively. For data transmission, the space-time encoding scheme with 2 transmit antennas and 1 or 2 receive antennas proposed in [2] is adopted. We use the space-time turbo equalization algorithm proposed in [14], and employ a rate 1/2 maximum free distance convolutional code [11] with constraint length 5 and generator polynomials (23, 35) in octal form. During each Monte-Carlo run, the block size is set to 2000 information bits followed by 4 tail bits to terminate the trellis, which corresponds to  $2004 \times 2 = 4008$  coded bits. They are interleaved by a random interleaver and transmitted over the SUI-3 MIMO channel. The turbo equalizer is investigated using both the channel estimated by the proposed channel estimator with orthogonal PVM sequences and the perfect channel state

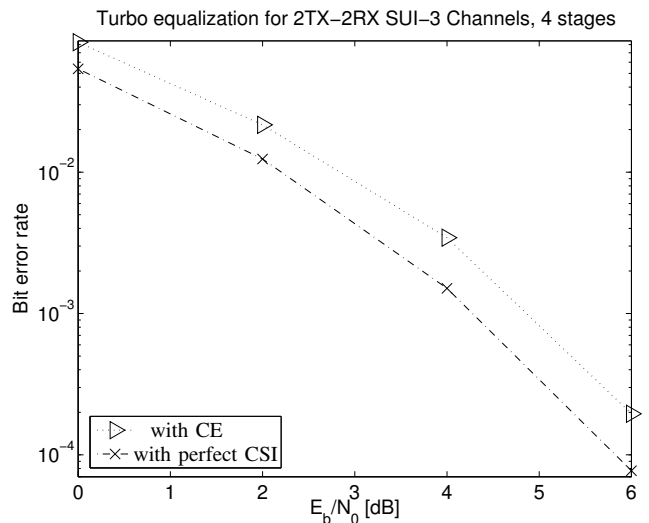


Fig. 6. Performance of the turbo equalization with channel estimate (CE) and perfect CSI for the 2TX-2RX system. The curves represent the 4th stage turbo equalization ( $D \times N = 12 \times 8$ ).

information (CSI). As shown in Fig. 5 and Fig. 6, the performance loss due to the channel estimation errors is within a fraction of 1 dB compared to the results achieved assuming perfect CSI. The gap is bigger for the 2TX-2RX system than the 2TX-1RX system. This is because the 2TX-2RX system can operate at a lower SNR while achieving comparable performance to that of the 2TX-1RX system, and so the estimation error increases as SNR decreases (as shown in Fig. 3). The performance loss can be effectively reduced by stacking more received samples (increasing  $D$ ) and/or by averaging more independent channel estimates (increasing  $N$ ). By comparing Fig. 5 with Fig. 6, we observed that the performance gain by adding one receiver antenna is approximately 6 dB<sup>2</sup> at BERs of  $10^{-3}$  and  $10^{-4}$ . We also learned from our experiments that the advantages of using multiple antennas is more significant for the coded system than for uncoded systems, especially when turbo equalization is used. These uncoded results are not shown here in order to conserve space.

## V. CONCLUSIONS

In this paper, we analyze a simple algorithm for estimating the frequency selective MIMO channels. Under the condition of orthogonal training sequences being transmitted, the proposed scheme is bound achieving and yields the same results as the optimum ML estimator. Some modifications of the proposed estimator with sample stacking and averaging have been discussed in this paper and shown to be very efficient for improving the performance of estimating static or quasi-static frequency-selective channels. The channel estimate is applied to a space-time turbo equalizer and is shown to yield close performance to that achieved assuming perfect CSI. Different training sequences are compared, and the results show that for the ML estimator, the choice of training sequences is not significant in the channel estimation performance; whereas

<sup>2</sup>Here  $E_b$  refers to the transmitted bit energy, and is not affected by the number of receiver antennas. The gain would be 3 dB if we define  $E_b$  as the received bit energy.

for the proposed estimator, the training sequences have to be orthogonal in order to achieve the desired results.

#### APPENDIX

Here, we derive the performance lower bound on the accuracy with which the channels can be estimated. Denote  $\mathbf{r} = \mathbf{S}\mathbf{h} + \mathbf{w}$  as the vector formed by  $D$  stacked samples of the received signal as shown in (6). The Cramer-Rao Lower Bound (CRLB) is a bound on the smallest covariance matrix that can be achieved by an unbiased estimator,  $\hat{\mathbf{h}}$ , of a parameter vector  $\mathbf{h}$ :

$$\mathbf{J}^{-1} \leq \mathbf{C}_{\hat{\mathbf{h}}} = \mathbf{E} \left\{ (\mathbf{h} - \hat{\mathbf{h}})(\mathbf{h} - \hat{\mathbf{h}})^* \right\};$$

$$\mathbf{J} = \mathbf{E} \left[ \left( \frac{\partial \ln p(\mathbf{r}; \mathbf{h})}{\partial \mathbf{h}} \right) \left( \frac{\partial \ln p(\mathbf{r}; \mathbf{h})}{\partial \mathbf{h}} \right)^* \right],$$

where  $\mathbf{J}$  is the Fisher information matrix and  $\ln p(\mathbf{r}; \mathbf{h})$  is the log-likelihood function of the observed vector  $\mathbf{r}$ . The vector  $\mathbf{r}$  is a complex Gaussian random vector, i.e.,  $\mathbf{r} \sim \mathcal{CN}(\mathbf{S}\mathbf{h}, N_0\mathbf{I})$  with likelihood function and log-likelihood function

$$p(\mathbf{r}; \mathbf{h}) = \frac{1}{(\pi N_0)^D} \exp \left[ -\frac{(\mathbf{r} - \mathbf{S}\mathbf{h})^*(\mathbf{r} - \mathbf{S}\mathbf{h})}{N_0} \right]$$

$$= \frac{1}{(\pi N_0)^D} \exp \left[ -\frac{\|\mathbf{r}\|^2 - \mathbf{h}^* \mathbf{S}^* \mathbf{r} - \mathbf{r}^* \mathbf{S} \mathbf{h} + \mathbf{h}^* \mathbf{S}^* \mathbf{S} \mathbf{h}}{N_0} \right];$$

$$\ln p(\mathbf{r}; \mathbf{h}) = \text{constant} - \frac{\|\mathbf{r}\|^2 - \mathbf{h}^* \mathbf{S}^* \mathbf{r} - \mathbf{r}^* \mathbf{S} \mathbf{h} + \mathbf{h}^* \mathbf{S}^* \mathbf{S} \mathbf{h}}{N_0}.$$

Taking the complex gradient [10] of  $\ln p(\mathbf{r}; \mathbf{h})$  with respect to  $\mathbf{h}$  yields

$$\frac{\partial \ln p(\mathbf{r}; \mathbf{h})}{\partial \mathbf{h}} = -\frac{1}{N_0} \frac{\partial [\|\mathbf{r}\|^2 - \mathbf{h}^* \mathbf{S}^* \mathbf{r} - \mathbf{r}^* \mathbf{S} \mathbf{h} + \mathbf{h}^* \mathbf{S}^* \mathbf{S} \mathbf{h}]}{\partial \mathbf{h}}$$

$$= -\frac{1}{N_0} (\mathbf{S}^* \mathbf{S} \mathbf{h} - \mathbf{S}^* \mathbf{r})^*.$$

The above equality holds since

$$\frac{\partial \|\mathbf{r}\|^2}{\partial \mathbf{h}} = \mathbf{0}; \quad \frac{\partial \mathbf{h}^* \mathbf{S}^* \mathbf{r}}{\partial \mathbf{h}} = \mathbf{0};$$

$$\frac{\partial \mathbf{r}^* \mathbf{S} \mathbf{h}}{\partial \mathbf{h}} = (\mathbf{S}^* \mathbf{r})^*; \quad \frac{\partial \mathbf{h}^* \mathbf{S}^* \mathbf{S} \mathbf{h}}{\partial \mathbf{h}} = (\mathbf{S}^* \mathbf{S} \mathbf{h})^*.$$

Thus we can derive,

$$\frac{\partial \ln p(\mathbf{r}; \mathbf{h})}{\partial \mathbf{h}^*} = \left( \frac{\partial \ln p(\mathbf{r}; \mathbf{h})}{\partial \mathbf{h}} \right)^* = \frac{\mathbf{S}^* \mathbf{r} - \mathbf{S}^* \mathbf{S} \mathbf{h}}{N_0}$$

$$= \frac{\mathbf{S}^* \mathbf{S}}{N_0} [(\mathbf{S}^* \mathbf{S})^{-1} \mathbf{S}^* \mathbf{r} - \mathbf{h}] = \mathbf{J}(\mathbf{h}) [\hat{\mathbf{h}} - \mathbf{h}].$$

This proves that the minimum variance unbiased estimator of  $\mathbf{h}$  is  $\hat{\mathbf{h}} = (\mathbf{S}^* \mathbf{S})^{-1} \mathbf{S}^* \mathbf{r}$ , which is equivalent to the ML estimator presented in Section III. It is efficient in that it attains the CRLB. The Fisher information matrix  $\mathbf{J}(\mathbf{h})$  and covariance matrix  $\mathbf{C}_{\hat{\mathbf{h}}}$  of this unbiased estimator are

$$\mathbf{J}(\mathbf{h}) = \mathbf{E} \left[ \frac{\mathbf{S}^* \mathbf{S}}{N_0} \right] \geq \frac{2E_b D \mathbf{I}_{2L}}{N_0};$$

$$\mathbf{C}_{\hat{\mathbf{h}}} = \mathbf{J}^{-1}(\mathbf{h}) = N_0 \mathbf{E} [(\mathbf{S}^* \mathbf{S})^{-1}] \geq \frac{N_0 \mathbf{I}_{2L}}{2E_b D}. \quad (16)$$

The above inequalities hold with equalities when columns of  $\mathbf{S}$  is orthogonal. The CRLB for the complex gain of each tap is each diagonal element of  $\mathbf{J}^{-1}(\mathbf{h})$ , i.e.,

$$\text{var}(\hat{h}_{ij}^l) = \text{diag}[\mathbf{C}_{\hat{\mathbf{h}}}]_{i \times L+j+l} \geq \frac{N_0}{2E_b D}. \quad (17)$$

#### ACKNOWLEDGEMENT

This work was sponsored by the UK Engineering and Physical Sciences Research Council under grant number EP/D07827X/1. The authors would like to acknowledge its financial support.

#### REFERENCES

- [1] V. Tarokh, N. Seshadri, A. Calderbank. "Space-time codes for high data rates wireless communications: performance criterion and code construction". *IEEE Transactions on Information Theory*, vol. 44, pp. 744–765, March 1998.
- [2] A. Alamouti. "A simple transmit diversity technique for wireless communications". *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.
- [3] V. Tarokh, N. Seshadri, A. Calderbank. "Space-time block coding from orthogonal designs". *IEEE Transactions on Information Theory*, vol. 45, pp. 1456–1467, July 1999.
- [4] V. Nguyen, L. White. "Joint space-time trellis decoding and channel estimation in correlated fading channels". *IEEE Signal Processing Letters*, vol. 11, no. 7, pp. 633–636, July 2004.
- [5] J. Choi. "Equalization and semi-blind channel estimation for space-time block coded signals over a frequency-selective fading channel". *IEEE Transactions on Signal Processing*, vol. 52, no. 3, pp. 774–785, March 2004.
- [6] H. Zhu, B. Farhang-Boroujeny, C. Schlegel. "Pilot embedding for joint channel estimation and data detection in MIMO communication systems". *IEEE Communications Letter*, vol. 7, no. 1, Jan. 2003.
- [7] S. Yang, J. Wu. "Optimal binary training sequence design for multiple-antenna systems over dispersive fading channels". *IEEE Transactions on Vehicular Technology*, vol. 51, no. 5, pp. 1271–1276, Sept. 2002.
- [8] IEEE 802.16 Working Group on Broadband Wireless Access Standards. available at <http://grouper.ieee.org/groups/802/16/>. 2002.
- [9] V. Erceg. "An empirically based path loss model for wireless channels in suburban environments". *IEEE JSAC*, vol. 17, no. 7, pp. 1205–1211, July 1999.
- [10] S. Kay. *Fundamentals of statistical signal processing*, Prentice Hall, NJ, 1998.
- [11] J. Proakis. *Digital Communications*, 3rd edition, McGraw-Hill, 1995.
- [12] P. Xiao, E. Ström. "Iterative demodulation of M-ary orthogonal signaling formats in coded DS-CDMA systems with soft interference cancellation and channel estimation". *IEICE Transactions on Communications*, vol. E89-B, no. 1, pp. 150–161, January 2006.
- [13] R. Iltis. "Joint estimation of PN code delay and multipath using the extended Kalman filter". *IEEE Transactions on Communications*, vol. 38, no. 10, pp. 1677–1685, Oct. 1990.
- [14] P. Xiao, R. Carrasco, I. Wassell. "Turbo equalization for space-time block coded FWA systems". *Proc. IEEE Information Theory Workshop 2006*, pp. 626–630, Oct. 2006.