Channel estimation in bit-interleaved coded modulation with iterative decoding

Z. Andalibi\textsuperscript{1,2} H.H. Nguyen\textsuperscript{1,2} J.E. Salt\textsuperscript{1,2}

\textsuperscript{1}Department of Electrical and Computer Engineering, University of Saskatchewan, 57 Campus Drive, Saskatoon, SK, Canada S7N 5A9
\textsuperscript{2}TRlabs, Saskatoon, Canada
E-mail: z.andalibi@usask.ca

Abstract: This study improves the fading channel estimation in bit-interleaved coded modulation systems with iterative decoding (BICM-ID) and signal space diversity (SSD) by embedding a training sequence. Existing training schemes work well at high signal-to-noise ratio (SNR) or slowly time-varying channels whereas the applications of BICM-ID are beneficial at low SNR and fast time-varying fading channels. Motivated by the power/bandwidth efficiency of the SSD technique and the fact that superimposed training outperforms pilot symbol-assisted modulation (PSAM) training over relatively fast time-varying channels, a new superimposed training sequence is explored. The proposed training sequence inserts pilot bits into the coded bits prior constellation mapping and signal rotation. This becomes a superimposed training sequence in the rotated symbols and helps the estimator to track fast variation of the channel gains. A soft iterative channel estimator is developed to work with the superimposed training sequence. The performance of the proposed scheme, namely SSD-pilot, is shown to be superior to PSAM scheme. To gauge the performance improvement achieved with the proposed channel estimation, an analytical bound on the asymptotic bit error probability for BICM-ID using SSD over correlated fading channels is provided. The Cramer–Rao bound on the mean-square error of the channel estimator is also derived to evaluate the performance of the iterative channel estimator.

1 Introduction

Signal space diversity (SSD) was introduced in [1, 2] as a power- and bandwidth-efficient technique for communication over fading channels. In SSD, an $N$-dimensional modulation scheme is created by partitioning the data into blocks of $N$ symbols and performing a rotation on each group of $N$-successive complex (two-dimensional) symbols. With such partitioning and rotation, the diversity order is maximised by increasing the minimum number of distinct components between any two $N$-dimensional constellation points. Among the various applications of SSD, bit-interleaved coded modulation (BICM) with iterative decoding (ID) over fast fading channels has been studied in [3]. In [3], the transmitter and receiver are designed under the assumption that the channel state information (CSI) is known at the receiver. However, in practical applications, this is not the case and the CSI must be estimated. Imperfect CSI due to channel estimation degrades the performance, especially at low signal-to-noise ratio (SNR) regime. This drawback can be mitigated by using a soft-iterative channel estimator [4].

Soft-iterative channel estimation has been intensively studied in turbo-coded systems, see for example, [5–7] and references therein. Soft-iterative channel estimation uses soft information from the soft-input soft-output (SISO) decoder in a semi-blind fashion to improve channel estimation performance, and consequently the system’s bit-error-rate (BER), over fading channels. The concept of semi-blind estimation is rooted in functions of both known and unknown signals. In the context of wireless communications, known signals can be either time-multiplexed or superimposed training sequence. In a time-multiplexed training sequence, pilot symbols are inserted regularly in the data stream. Such schemes, which are commonly called pilot-symbol assisted modulation (PSAM) schemes perform well over slowly time-varying channels [8, 9]. To work in a fast fading channel, the pilot
symbols in a PSAM scheme have to be inserted more often, which causes bandwidth expansion. Alternatively, a superimposed training scheme, which in effect superimposes a training signal on each transmitted symbol, can be used. Superimposed training schemes work well in fast fading channels provided that the SNR is sufficiently high. Such schemes are also effective in multi-path channels [10].

Although there has been no study on the impact of channel estimation on the design of the transmitter and receiver for a BICM-ID-SSD system, the use of PSAM or superimposed schemes was widely considered for coded and un-coded systems, see, for example [11–17] (Hereafter the abbreviation BICM-ID-SSD refers to a system that implements bit-interleaved coded modulation, iterative decoding and signal space diversity technique.). In particular, using pilot symbols for channel estimation in turbo-coded systems has been examined in [11]. The study assumed that the channel is flat fading with a gain that depends on time but is constant for the duration of a symbol. Such channels are referred to as correlated fading channels. The study in [11] used an iteratively filtered PSAM algorithm in the receiver. Reference [12] proposes sparsely interleaved estimation and decoding as an alternative algorithm for iterative channel estimation in BICM-ID systems. In these papers, pilot symbols are used to initialise the channel estimates and then these estimates are refined in subsequent iterations by an interpolator that uses information fed back from the decoder. In [12] it was shown that in order to improve the performance, the refining phase of the channel estimation is not needed for each iteration of decoding stage. In [13], the performances of PSAM and superimposed training schemes are compared using the worst-case un-coded BER as the performance measure. It was shown for uncoded systems at high SNR that PSAM performs better for slowly time-varying fading channels and the superimposed training scheme performs better in fast time-varying fading channels. This observation also applies to coded systems. For BICM systems, the performance of PSAM has been studied in [14] for a range of SNR. In particular, the relation between fade rates and required power of pilots has been investigated using extrinsic information transfer chart. Similar analysis has been investigated for coded systems with superimposed schemes in [15] using BER simulation. However, at low SNR in a fast time-varying fading environment, conventional training schemes fail. On the other hand, since BICM works well at low SNR and over a fast time-varying channel, it is of interest to develop a new training scheme for this system.

This paper proposes and examines alternative training sequences for a BICM-ID-SSD system operating over a correlated fading channel. Specifically, the paper investigates injecting pilots at the bit level instead of the symbol level. The intent is to exploit SSD technique to improve the performance of the iterative receiver. The proposed method inserts pilot bits into the coded bits before constellation mapping and signal rotation. As a result, the rotated superimposed symbols, which shall be referred to as SSD-pilots, can be considered as a superimposed training sequence that is inherent to the rotated transmitted symbols. A soft iterative receiver is developed for the bit level embedded training sequence. In particular, the estimator in the proposed receiver uses the training sequence in addition to the soft information from the decoder to produce estimates of the time-varying channel gains for the demodulator. Moreover, the demodulator uses the training part in the received signal to provide reliable extrinsic information for the decoder. It shall be demonstrated that the performance of the proposed receiver is significantly better than that of the conventional PSAM scheme.

The paper is organised as follows. The system model of BICM-ID-SSD is presented in Section 2. In this section, the structure of the training sequence is described and the transmitter and receiver are developed for the proposed method of training design. Section 3 analyses the performance of the BICM-ID-SSD system over a correlated fading channel. In particular, a lower bound of the asymptotic BER is first obtained under the perfect CSI. Then the Cramer–Rao bound (CRB) for the proposed channel estimation method is derived. Section 4 provides numerical results and performance comparisons. Section 5 offers conclusion.

Notation: Upper and lower boldface letters denote matrices and column vectors, respectively. Superscripts $(\cdot)^H$ and $(\cdot)^T$ indicate Hermitian and transpose, respectively. For matrix $A$, $A_{i,j}$ denotes its $(i,j)$th entry while $A_j$ indicates the $j$th row. The log-likelihood ratio of bit $b$ is defined as $\Lambda^{(b)} = \log[P(b=1)/P(b=-1)].$

2 System model

The block diagram of the transmitter and receiver in discrete-time equivalent baseband is shown in Fig. 1.

2.1 Transmitter

A sequence $\{u_j\}$, $1 \leq j \leq L$, of information bits is first encoded by a rate-$r$ convolutional encoder. The encoded bits $\{e_i\}$, $1 \leq i \leq L/r$ are then passed to a bit-interleaver with a length of $L/r$. Next, the interleaved sequence $\{i_j\}$ is segmented into groups of $(N - N_p) \times m$ bits, where $N$ is the size of the rotation matrix, $N_p$ is the number of pilot symbols in $N$ rotated symbols and $m$ is the number of bits carried by one symbol of a QAM constellation whose size is $|\Omega| = 2^n$. Next, known pilot bits $\{p_n\}$, $n = 1, 2, \ldots, mN_p$, is inserted in the segmented group of $(N - N_p) \times m$ bits. The pilot bits can be placed at the beginning, at the end or somewhere in between as long as every $m$ pilot bits are inserted as a block in the segmented group and mapped to one QAM
constellation point. This is necessary so that the channel estimator can work with known pilot QAM symbols in the initialisation phase. Each expanded group of size $N m$ is then mapped to one complex $N$-dimensional ‘super’ symbol, $s = [s_1, s_2, \ldots, s_N]^T$.

There are many possible mappings between $N m$ bits and a super symbol. The sigma mapping scheme in [3] facilitates a low-complexity soft-output minimum mean-square error (MMSE) demodulator and is used here for the same reason. In general, the sigma mapping relates a binary vector $z$ to the super symbol $s$ with the equation

$$s = V(2z - 1) = Vb$$  \hspace{1cm} (1)$$

where $V$, $z$ and $b$ are defined as follows:

- $V = \text{diag}(\psi, \psi, \ldots, \psi)$ is an $N \times N m$ block diagonal matrix and $\psi = [\psi_1, \psi_2, \ldots, \psi_m]$ is the basis vector whose elements are unit-norm complex numbers.

- $z$ is an $N m \times 1$ binary vector of interleaved coded bits and pilot bits whose entries are either 1 or 0.

- $b$ is a binary vector whose entries are $\pm 1$ and represents both pilot and coded bits.

With the proper design of $\psi$, each component $s_i$ is guaranteed to belong to a standard two-dimensional QAM constellation $\Omega$ [3]. As an example, Fig. 2 shows the sigma mapping of 4-QAM and 16-QAM.

Rotation is accomplished with a matrix, $G$, which could take on one of several forms. One form that can be used

$$G = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 & \alpha_1 & \cdots & \alpha_{N-1}^N \\ 1 & \alpha_2 & \cdots & \alpha_{N-1}^N \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_N & \cdots & \alpha_{N-1}^N \end{pmatrix}$$  \hspace{1cm} (2)$$

where $\alpha_i = \exp(j[2\pi/4N])\exp(j[2\pi(i - 1)/N])$ and $j = \sqrt{-1}$. The matrix rotates a super symbol $s$ into a rotated super symbol, denoted by $x = Gs$.

Let $\Gamma = GV$ and introduce subscripts as the time index to the super symbols and rotated super symbols. Then the $\ell$th rotated super symbol can be represented by

$$x_\ell = [x_{\ell-1}N+1, \ldots, x_{\ell-1}N+N]^T = \Gamma b_\ell, \quad 1 \leq \ell \leq \frac{L}{N - N_p m}$$  \hspace{1cm} (3)$$
Thus by showing the time index in the form of 
\[ k = (l - 1)N + i, \quad 1 \leq i \leq N, \] 
the \( k \)th transmitted complex symbol is given by
\[ x_k = \Gamma_k b_k = x_k^{(d)} + x_k^{(p)}, \quad k = 1, 2, \ldots, \frac{N}{N - N_p rm} \]
where \( x_k^{(d)} \) and \( x_k^{(p)} \) are the data and pilot portions of the transmitted symbols, respectively, and \( l = [k/N] \) and \( i = k - (l - 1)N \). For the ease of exposition, it is assumed that all the pilot bits take on the same value of 1.

It is of interest to measure the effective energy of the system. The energy efficiency of SSD-pilot scheme can be expressed as
\[ \epsilon_{SSD-pilot} = \frac{(N - N_p)E_i}{(N - N_p)E_i + N_p E_p} = \frac{N - N_p}{N - N_p + \eta N_p} \]
when \( E_p = \eta E_i \) and \( E_i \) and \( E_c \) are the energy per pilot symbol and the energy per data symbol before rotation, respectively. For the purpose of comparison to PSAM, it is pointed out that the energy efficiency of PSAM is given by \( 1/(M - 1/\eta) \), where \( M \) is the pilot symbol spacing.

### 2.2 Iterative receiver

The sequence of received symbols, denoted \( \{r_k\} \), is given by
\[ r_k = b_k x_k + w_k, \quad k = 1, 2, \ldots, \frac{N}{N - N_p rm} \]
where \( w_k \) is a noise sample and \( b_k \) is a channel gain at time \( k \).

The sequence of \( \{w_k\} \) is a set of i.i.d. zero-mean circularly symmetric Gaussian random variables with variance \( N_0 \).

The sequence \( \{b_k\} \) contains complex Gaussian random variables with zero mean and autocorrelation \( R(k) = \sigma_b^2 f_d(2\pi f_d T_k) \), which is given by the Jakes’s model.

An iterative receiver operates on the received signals in an iterative manner. In the first iteration, the received signals are filtered to produce initial estimates of the channel gains.

In this section, an iterative MMSE channel estimator is developed for the SSD-pilot training sequence. It is shown in Section 4 that inserting pilot bits with the coded bits lowers error probability in each successive iteration.

### 2.3 Channel estimator

In subsequent iterations, soft information from the decoder is used to improve the performance of the channel estimator. The channel estimator uses such information to compute new estimates of the channel coefficients using expected values of the data symbols.

By using (4), \( E[x_k] \) and \( \sigma_{x_k}^2 \) are computed as
\[ E[x_k] = \Gamma_k E(b_k) \]
\[ \sigma_{x_k}^2 = \frac{\sum_{j=1}^{N_m} |\Gamma_{ij}|^2 \sigma_{\hat{\beta}(\hat{\alpha})+1+j}^2}{N} \]

where the \( \{\Lambda_{\hat{\alpha}}^{(p)}(\hat{\alpha})\} \) should be used to calculate \( E(b_k) = \tan(b_{\hat{\alpha}})/2 \) and \( \sigma_{x_k}^2 = 1 - |E(b_k)|^2 \) for the pilot bits, since \( \{\rho_k\} \) are known at the receiver, one has \( E(b_k) = b_k \).

The demodulator uses the \( \{\Lambda_{\hat{\alpha}}^{(d)}(\hat{\alpha})\} \) and \( \Lambda_{\hat{\alpha}}^{(d)}(\hat{\alpha}) \) directly to calculate \( \{\Lambda_{\hat{\alpha}}^{(p)}(\hat{\alpha})\} \) for the rest of the iterations, as in the first iteration. The detailed operation of the channel estimator is discussed next.
• The pilot portion of the transmitted symbols is \( \tilde{x}_k^{(p)} = [x_{k-K_0}^{(p)}, \ldots, x_k^{(p)}, \ldots, x_{k+K_0}^{(p)}]^T \), which is a known vector at the receiver.

• The vector of the transmitted symbols, \( \tilde{x}_k = [x_{k-K_0}, \ldots, x_k, \ldots, x_{k+K_0}]^T \) is the collection of the \( (2K_0 + 1) \) symbols that surround \( x_k \).

• The covariance matrix of the observation vector is obtained with \( E[\tilde{r}_k\tilde{r}_k^H] \).

• The covariance vector, which has length \( (2K_0 + 1) \) is obtained by \( E[\tilde{r}_k\tilde{r}_k^H] \).

With these definitions, the linear MMSE estimate of the channel gain is

\[
\hat{g}_k = (E[\tilde{r}_k\tilde{r}_k^H])^{-1} E[\tilde{r}_k\tilde{r}_k^H]^H \tilde{r}_k \tag{9}
\]

Expressions for the covariance matrix and covariance vector must be obtained to calculate \( \hat{g}_k \) with (9) for the first iteration, where only the pilot portions of the received signals are known. For the pilot portion \( E[x_k^{(p)}(x_j^{(p)})^*] = \delta_k^j \delta_k^j \), \( \forall i,j \) and \( E[x_k^{(p)}(x_j^{(p)})^*] = \delta_k^j \). For the data portion of the transmitted signal, \( E[x_k^{(p)}(x_j^{(p)})^*] = ((N - N_p)/N)E_k \) for \( i = j \), \( E[x_k^{(p)}(x_j^{(p)})^*] = 0 \) for \( i \neq j \) and \( E[x_k^{(p)}] = 0 \) for the first iteration. Assembling this information leads to

\[
E[\tilde{r}_k\tilde{r}_k^H]_{i,j} = \begin{cases} \left( \delta_k^j (\frac{N - N_p}{N}) + \frac{N}{N} \right) \delta_k^j + N_0, & i = j \\ \delta_k^j (\delta_k^j)^* \sigma_h^2 J_0(2\pi f_d|i-j|T), & i \neq j \end{cases}
\] \tag{10}

where \( E[\tilde{r}_k\tilde{r}_k^H]_{i,j} \) is the \((i,j)\)th element of the covariance matrix and \( \tilde{x}_k^{(p)} = x_k^{(p)} + (K_0+1) \) is the \( i \)th element of vector \( \tilde{x}_k^{(p)} \). In addition, \( E[\tilde{r}_k\tilde{r}_k^H] = \sigma_h^2 J_0(2\pi f_d|i-j|T) \delta_k^j \). Note that the above calculations imply that the matrix \( E[\tilde{r}_k\tilde{r}_k^H] \) and vector \( E[\tilde{r}_k\tilde{r}_k^H] \) can be calculated off-line for the first iteration.

In the next iterations, for calculating the covariance matrix and covariance vector, the soft information of data is taken into account. Using (7), \( E[x_k(x_j)^*] = |E[x_k]|^2 + \sigma_h^2 \) if \( i = j \) and \( E[x_k(x_j)^*] = E[x_k]E[x_j]^* \) for \( i \neq j \). Then the elements of the covariance matrix are

\[
E[\tilde{r}_k\tilde{r}_k^H]_{i,j} = \begin{cases} \left( |E[x_k]|^2 + \sigma_h^2 \right) \delta_k^j + N_0, & i = j \\ E[x_k(x_j)^*] \delta_k^j J_0(2\pi f_d|i-j|T), & i \neq j \end{cases}
\] \tag{11}

and the entries of the covariance vector are obtained as \( E[\tilde{r}_k\tilde{r}_k^H] \), \( \delta_k^j J_0(2\pi f_d|i-j|T) \times E[x_k] \). In the above last two expressions \( E[x_k(x_j)^*] \) and \( \sigma_h^2 \) are computed using (7) and (8), respectively, and by associating \( \tilde{x}_k \) to the \( i \)th element of vector \( \tilde{x}_k \) that is, \( \tilde{x}_k = x_k^{(p)} + (K_0+1) \).

3 Bound for asymptotic error performance and Cramer–Rao bound for channel estimation

3.1 Analytical bound of asymptotic BER with perfect CSI

The union bound on the BER for BICM-ID-SSD is helpful for evaluating the system performance in a correlated fading channel. The bound for a rate-\( r \) convolutional code can be written as

\[
P_b \leq \frac{1}{r} \sum_{d=d_1}^{m} c_d f(d, \Psi, \zeta) \tag{12}
\]

where parameters \( d, d_1, \Psi, \zeta \) and \( \zeta \) as well as function \( f(\cdot) \) are defined as follows: \( d \) is the Hamming distance, \( d_1 \) is the free Hamming distance of the code, \( c_d \) is the total input weight of error events at Hamming distance \( d \), \( \Psi \) is the \( N \)-dimensional complex constellation set and \( \zeta \) is the mapping rule. The function \( f(d, \Psi, \zeta) \) denotes the average pairwise error probability (PEP), which depends on the Hamming distance \( d \), the signal constellation set \( \Psi \) and the mapping rule \( \zeta \). In general, \( f(d, \Psi, \zeta) \) can be approximated as in [3] by

\[
f(d, \Psi, \zeta) \approx \frac{1}{2} (\delta(G, \Omega, \zeta))^d \tag{13}
\]

where, for an uncorrelated Rayleigh fading channel

\[
\delta(G, \Omega, \zeta) = \frac{1}{N} \frac{1}{m \omega^m} \sum_{\omega=1}^{N} \left[ \sum_{i=1}^{m} \sum_{j=1}^{N} \prod_{k=1}^{N} \left( 1 + \frac{|G_{\omega}(s_k - s_{k'-j})|^2}{4N_0} \right)^{-1} \right] \tag{14}
\]

In the above \( s_k \) is a constellation point in \( \Omega \) and notation \( k \sim j \) is used to refer to a signal whose mapping is the same as that of \( s_k \), except that the \( j \)th bit is toggled (i.e. complemented). As an example for 16-QAM, if \( s_k \) has the mapping 1111 and \( j = 2 \) then \( s_{k-j} \) would be the signal that has the mapping 1110. As an illustration, the signals \( s_{15} \) and \( s_{15-2} \) are shown in Fig. 2 for the sigma mapping scheme.

The union bound over a correlated fading channel is now derived using the procedure of [3]. To account for the correlation, the joint probability density function of the channel is expressed as

\[
\rho(h) = \frac{1}{\pi \det C} \exp(-h^H C^{-1} h) \tag{15}
\]

where \( b = [b_1, \ldots, b_N]^T \) is a correlated fading vector and
\( C = E[\mathbf{bb}^\dagger] \) is an \( N \times N \) covariance matrix, with

\[
C_{i,j} = \begin{cases} \sigma_s^2, & i = j \\ \sigma_s^2 J_0(2\pi f_d(i-j)T_s), & i \neq j \end{cases}
\]

(16)

After some manipulation and then diagonalising the data vector, that is, \( \mathbf{D} = \text{diag}(|G_{1,1} s_1|^2, \ldots, |G_{N,L} s_L|^2) \), and using the method in [21], one recognises that \( f(d, \Psi, \xi) \) has the same form as (13), but with \( \delta(G, \Omega, \xi) \) given as

\[
\delta(G, \Omega, \xi) = \frac{1}{N} \frac{1}{m} \sum_{i=1}^{N} \left[ \sum_{j=1}^{m} \prod_{j=1}^{N} \left( 1 + \frac{\lambda_{i,j}^2}{4N_0} \right)^{-1} \right]
\]

(17)

where \( \lambda_{i,j} \) are the eigenvalues of \( \mathbf{DC} \). It is clear that there are at most \( N \) non-zero eigenvalues. It is also noted that the number of eigenvalues equals the average number of distinct channel states in one rotated super symbol. Therefore it follows from (13) that the diversity order of the system must be less than or equal to \( N_{\text{CI}} \). If the rank of \( \mathbf{DC} \) is less than \( N_{\text{CI}} \), then the number of non-zero eigenvalues and therefore the diversity order is less than or equal to the rank of \( \mathbf{C} \) times \( d_{\text{CI}} \). For example, if \( f_d T_s = 0.02 \), the rank of \( \mathbf{C} \) is 8 when \( N = 32 \). This implies that the maximum diversity order is \( 8d_{\text{CI}} \).

### 3.2 CRB for mean-square error of the channel estimation

CRB is widely used as a benchmark for the performance of any channel estimator. The bound states that the mean-square error (MSE) matrix of any [unbiased] estimator \( \mathbf{b} \) is lower bounded by [22]

\[
\text{MSE}(\mathbf{b}) = E[(\mathbf{b} - \hat{\mathbf{b}})(\mathbf{b} - \hat{\mathbf{b}})^\dagger] \geq J(\mathbf{b})^{-1}
\]

where \( J(\mathbf{b}) \) is the complex Fisher information matrix defined by

\[
J(\mathbf{b}) = E\left\{ \left[ \frac{\partial \ln \rho(r, \mathbf{b})}{\partial \mathbf{b}} \right] \left[ \frac{\partial \ln \rho(r, \mathbf{b})}{\partial \mathbf{b}} \right]^\dagger \right\}
\]

(18)

A closed-form expression for \( J(\mathbf{b}) \) is found by taking the expectation of (18) over \( r \) and \( \mathbf{b} \). To this end, it is convenient to express \( \rho(r, \mathbf{b}) \) as \( \rho(\mathbf{b}) \cdot \rho(\mathbf{b})^\dagger \), where \( \rho(\mathbf{b}) \) is defined in (15), \( \rho(\mathbf{b}) = [1/(\pi N_0)^{N/2}] \exp(-1/(2N_0)\|r - X_D^\dagger \mathbf{b}\|^2) \), \( r = [r_1, \ldots, r_N]^\dagger = X_D^\dagger \mathbf{b} + \mathbf{w} \) and \( X_D = \text{diag}(x_1, \ldots, x_N) \). After some manipulations, the Fisher information matrix is given in terms of \( X_D \) as follows

\[
J(\mathbf{b}) = \frac{1}{N_0} E\{X_D X_D^\dagger\} + C^{-1}
\]

(19)

Note that \( E\{X_D X_D^\dagger\} \) is a diagonal matrix which has \( E\{x_i^2\} = |E[x_i]|^2 + \sigma_s^2 \) on its diagonal, \( i = 1, \ldots, N \). The terms \( E[x_i] \) and \( \sigma_s^2 \) can be numerically calculated based on (7) and (8). The vast simulations are arranged even at low SNR to obtain an accurate CRB.

### 4 Simulation results

The performance of the proposed method is investigated using a computer simulation to calculate the MSE of the channel estimator and the resulting BER. The same convolutional code, data frame, bit-interleaver and modulation format are used in all simulations. The channel code is a rate-1/2 convolutional code with constraint length 5 and generator polynomial \( [1 + D + D^3 + D^4; 1 + D + D^4] \). The data frame consists of 11 996 data bits and four tail bits. The bit-interleaver has length \( L = 2 \times 12 000 \). The modulation is quadrature phase-shift keying with sigma mapping, where the basis vector is \( \mathbf{v} = [1, \mathbf{l}] \).

Fig. 3 plots BER against \( E_s/N_0 \), where \( E_s \) is the energy per information bit, and the normalised fade rate is taken to be \( f_d T_s = 0.02 \). The size of the rotation matrix is \( N = 16 \) whereas \( N_{\text{CI}} = 1 \) and \( \eta = 1 \). From (5) this means that the energy efficiency is \( \xi = 0.9375 \), since \( E_s = E_p = 2 \). The length of the Wiener filter is set to \( 2K_0 + 1 \) with \( K_0 = 40 \). This has the Wiener filter estimating each channel tap with 81 channel realisations. For each value of \( E_s/N_0 \) considered, \( 10^5 \) independent frames were run.

The performance of the SSD-pilot scheme is compared with the performance of the PSAM training scheme as well as the analytical bound in Fig. 3. Note that for a fair comparison, the channel estimator used with PSAM is also based on the iterative approach with rotated data symbols. For the PSAM scheme, the pilot symbol is \( 1 + j \) and the pilot spacing is \( M = 16 \), which yields the same energy efficiency as that of SSD-pilot scheme. The other simulation parameters are the same as in SSD-pilot scheme.

**Figure 3** Comparison of BER obtained from SSD-pilot and PSAM with the iterative channel estimator when \( N = 16 \), \( f_d T_s = 0.02 \) and \( N_p = 1 \), and for one and five iterations
There are five curves in Fig. 3. The top two provide the performance after the first iteration for the SSD-pilot and the PSAM schemes. The middle two curves show the performance after five iterations. The bottom curve, shown as a dashed line, is a plot of the bound. The curves for the first iteration show that the performance of the proposed scheme is about 2 dB worse than the conventional PSAM scheme. This is expected because only the pilot portions of the transmitted symbols are used for channel estimation in the first iteration in SSD-pilot scheme. However, the results are quite different after five iterations. The SSD-pilot scheme is from 1 to 1.5 dB better than the PSAM scheme, depending on $E_b/N_0$. The reason for this is that the pilot information is embedded in the rotated symbols for the SSD-pilot scheme and not for the PSAM scheme. Since the SSD-pilots are embedded in the SSD received symbols, the demodulator can make use of this information for its training. In contrast, the pilots in PSAM, which are not embedded in the rotated symbols, cannot be used by the demodulator. In the end the SSD-pilot information used in the demodulator iteratively improves the overall performance of the receiver. The improvement is such that after five iterations the performance closely approaches the analytical bound. Note that there is a gap between the union bounds and the BER curves, which is because the union bound is calculated assuming perfect CSI.

Fig. 4 has the same curves as Fig. 3 for a fade rate of $f_d T_s = 0.05$. For this larger fade rate, parameters $N$, $N_p$, and $K_0$ are changed to accommodate better channel tracking. The following values were used: $N = 8$, $N_p = 1$, and $K_0 = 45$. For the PSAM scheme, $M$ was changed from 16 to 8 to make the energy efficiency the same. It is observed that the SSD-pilot method also outperforms conventional PSAM under this faster fade rate. SSD-pilot gain is as high as 1.25 dB at lower BERs. From these two figures it can be seen that the tracking ability of SSD-pilot is better than PSAM scheme.

The MSE performance of the SSD-pilot estimator is shown in Figs. 5 and 6 after one, two, five and eight iterations for $f_d T_s = 0.02$ and $f_d T_s = 0.05$, respectively. The CRBs of the channel estimator are also plotted in Figs. 5 and 6 (dashed lines) for comparison. The simulation results after five iterations are very close to the CRBs in both figures. It is observed that MSE curves for $E_b/N_0 \geq 5$ dB exhibit flat slopes with higher numbers of iterations. This is explained as follows. In high $E_b/N_0$ region, the genie condition (i.e. error-free feedback from the decoder) is basically achieved after a sufficiently large number of iterations (say five iterations). This condition implies that,

![Figure 4](image)

**Figure 4** Comparison of BER obtained from SSD-pilot and PSAM with the iterative channel estimator when $N = 8$, $f_d T_s = 0.05$ and $N_p = 1$, and for one and five iterations

![Figure 5](image)

**Figure 5** MSE obtained from SSD-pilot with the iterative channel estimator when $N = 16$, $f_d T_s = 0.02$ and $N_p = 1$, and for one, two, five and eight iterations

![Figure 6](image)

**Figure 6** MSE obtained from SSD-pilot with the iterative channel estimator when $N = 8$, $f_d T_s = 0.05$ and $N_p = 1$, and for one, two, five and eight iterations
The soft iterative channel estimator was developed to exploit the pilot information and the soft information of the rotated symbols. In order to analyse the effectiveness of using SSD-pilot symbols, the analytical bound of the BER over a correlated channel with perfect CSI was obtained. Simulation results verified that, compared to PSAM, the BER performance is improved by about 1 dB at the BER level of $10^{-4}$ under both fade rates of 0.02 and 0.05 and at the same energy efficiency. Finally, a CRB for the MSE of the estimated channel coefficients was derived. Simulation results showed that the MSE of the iterative channel estimator closely approaches the CRB after five iterations.

6 References


