RESEARCH

Open Access

Training design for precoded BICM-MIMO systems in block-fading channels

Zohreh Andalibi[1,2*], Ha H Nguyen[1,2] and Joseph E Salt[1,2]

Abstract

In order to improve bandwidth efficiency and error performance, a new training scheme is proposed for bit-interleaved-coded modulation in multiple-input multiple-output (BICM-MIMO) systems. Typically, in a block-fading channel, the training overhead used for obtaining channel knowledge is proportional to a power of 2 of the number of transmit antennas. However, this overhead can be reduced by embedding pilot symbols within data symbols before precoding. The values, positions, and the number of pilot symbols are found by minimizing the Cramer-Rao bound on the channel estimation error. Computer simulations are presented to demonstrate the advantage of the proposed scheme over other training methods, in terms of both the mean-square-error of the channel estimation and the system's frame-error-rate.

Keywords: BICM-MIMO, block fading, channel estimation, training design, pilot symbols, Cramer-Rao bound, iterative receiver

1 Introduction

The pioneering work on multiple-input multiple-output (MIMO) systems [1] shows that a MIMO system can provide a multiplexing gain and accordingly high spectral efficiency over slow fading channels. On the other hand, to achieve a high diversity order, space-time transmission techniques can be implemented at the transmitter [2,3]. To achieve both high diversity order and coding gain in coded modulation systems, the concept of space-time transmission has also been applied [4,5]. In such systems, space-time transmission is typically implemented using a linear space-time matrix, or equivalently a linear precoder, so that a single modulation symbol is efficiently transmitted across multiple transmit antennas. Among many research works on precoder design for coded modulation systems with multiple antennas, the design that considers all the relevant components of the transmitter, namely precoding, modulation, and interleaver, can be found in [5-7]. Specifically, a *full-rate* precoder with any size and for any number of transmit antennas is designed in [6] to maximize the achievable diversity order and coding gain in MIMO block-fading channels.

It is shown in [6] that the maximum achievable diversity order can be realized by an iterative receiver that employs a soft-input soft-output detector [5] and under the assumption of having the *perfect* channel state information (CSI) at the receiver. In practice, however, CSI has to be estimated using a channel estimator and it is never perfect. Two types of channel estimators have been used for MIMO block-fading channels in coded modulation systems, i.e., training-based and semi-blind channel estimators [8,9]. In both types of channel estimators, known signals are used to estimate the CSI at the first iteration of the iterative receiver.

Conventionally, for block-fading channels, known signals or the training sequence is included at the beginning of each data block, which is called time-multiplexed training or pilot symbol-assisted modulation (PSAM) scheme [10]. This scheme however reduces bandwidth efficiency of MIMO systems, since the amount of training overhead needed is at least a power of 2 of the number of transmit antennas [11] to ensure the identifiability of the MIMO channel. A straightforward application of the PSAM scheme to a BICM-MIMO system would be time-multiplex data information with the training information after the precoder.

As an alternative to the above conventional PSAM scheme, a potential benefit can be sought by time-

* Correspondence: z.andalibi@usask.ca
[1]TRLabs, Saskatoon, Canada
Full list of author information is available at the end of the article

Springer

multiplexing data information with the training information *before* the precoder in the transmitter. This new approach shall reduce the required training overhead compared to the conventional PSAM, since the transmitted training symbols are spread over more time periods; thanks to the precoder. This approach shall be referred to as precoded PSAM (PPSAM). Investigating power and time allocations of the training symbols in PPSAM scheme is the main objective of this article.

Moreover, by multiplexing the training sequence before precoder, training symbols can be exploited in both the initialization and iteration phases of the iterative channel estimation process. This is different from a conventional iterative channel estimator using PSAM scheme, in which training sequence is only used at the initialization phase. A natural question is whether the optimal training design for the initialization phase using PPSAM scheme is still optimal for subsequent iterations of an iterative channel estimator. On the one hand, the channel estimation error at the initialization phase translates to an SNR shift in the BER performance [8]. On the other hand, the channel estimation error from the last iteration of the iterative estimator has a strong impact on the error floor of the BER performance [12]. Therefore, optimal training sequence should be designed carefully that considers both initialization and iteration phases.

One of different criteria that have been used to design training sequences is the minimization of the Cramer-Rao bound (CRB) of the channel estimation error [10]. This criterion shall be used in this article due to two main reasons. First, it is directly related to the channel estimation error. Second, since the CRB is a lower bound on the mean-squared-error (MSE) of any unbiased estimator, designing training sequences using this criterion would be applicable to many estimation algorithms. Other design criteria, such as maximizing the channel capacity [8] and minimizing the outage probability [13], are based on some specific channel estimation algorithms.

The article is organized as follows. The system model of BICM-MIMO is presented in Section 2. In Section 3 a lower bound on the MSE of the channel estimator is obtained and the training sequence is designed by minimizing this bound. Section 4 provides numerical results and comparisons. Section 5 concludes the article.

2 System model
Figure 1 shows the block diagram of a BICM-MIMO system under consideration. At the transmitter, a channel encoder with a rate-r error-correcting code converts the vector of information bits $\mathbf{b}$ into a codeword $\mathbf{c}$. The coded bits are then interleaved by a random interleaver as described in [6] to produce the interleaved codeword

$\tilde{\mathbf{c}}$. The interleaved codewords are segmented into groups of $(Nn_t - N_p) \times m$ bits, where N is the spreading factor of the precoder, n_t is the number of transmit antennas, N_p is the number of pilot symbols in Nn_t precoded symbols and m is the number of bits carried by one symbol of a QAM constellation whose size is $|\Omega| = 2^m$. Next, the coded bits are mapped to $(Nn_t - N_p)$ QAM constellation points. In this step, N_p known pilot symbols are inserted in every segmented group of $(Nn_t - N_p)$ data symbols to produce N super-symbols. Here, each super-symbol refers to a group of n_t consecutive symbols. Investigating the positions and the number of pilot symbols (i.e., N_p) to be used in each Nn_t symbols is the main objective of this article.

Every group of N super-symbols is then spread over N time periods using a linear precoder $\mathbf{G}$. The $Nn_t \times Nn_t$ matrix $\mathbf{G}$ multiplies a vector of Nn_t QAM symbols at the precoder input, and generates Nn_t symbols to be transmitted over $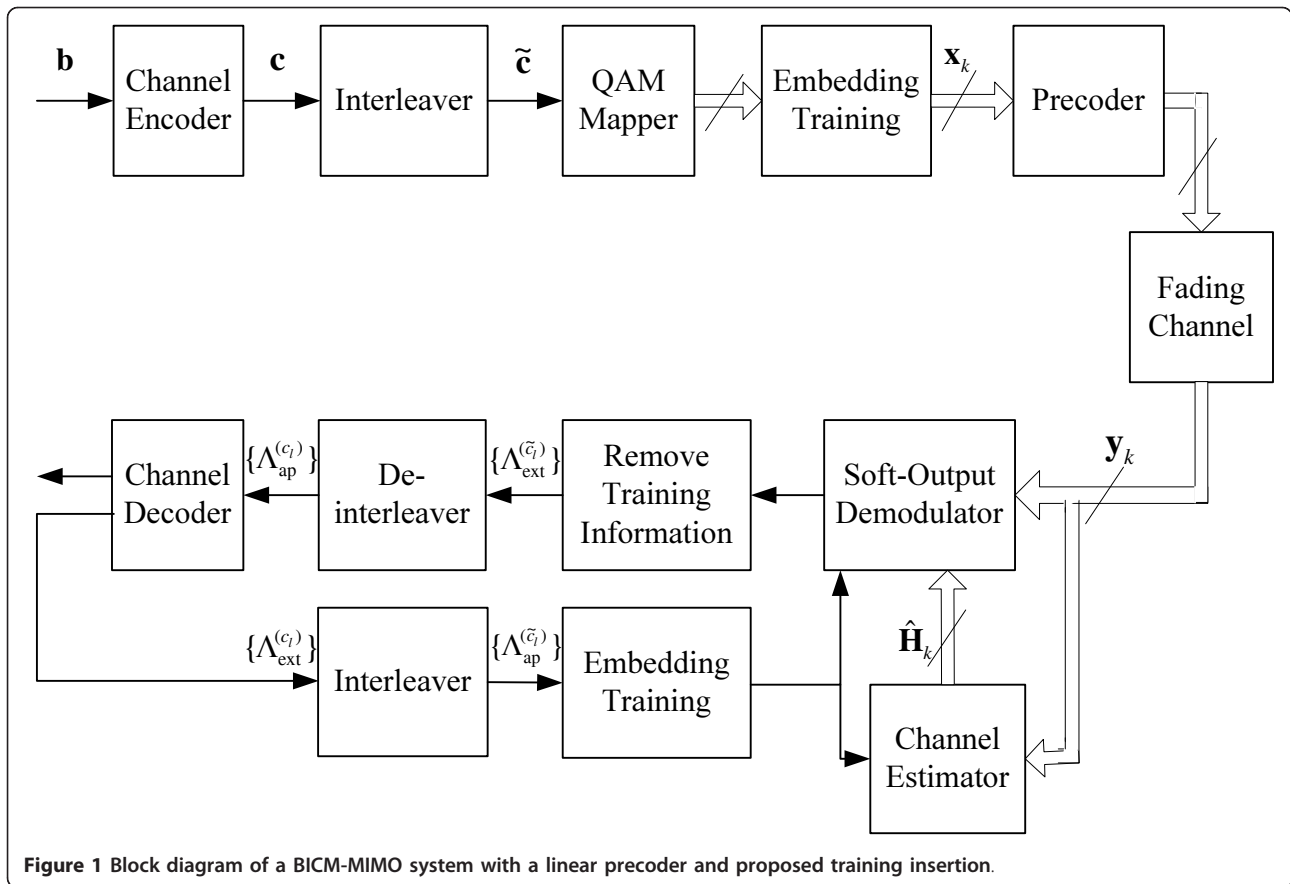n_t$ antennas, over N time periods. This is illustrated in Figure 2. Let $\mathbf{x}_k = [x_{(k-1)Nn_t+1}, x_{(k-1)Nn_t+2}, \ldots, x_{(k-1)Nn_t+Nn_t}]$ be the kth vector to be precoded. Then, $\mathbf{x}_k\mathbf{G}$ gives the precoded symbols. Here, x_i's are complex data or pilot symbols belonging to the 2^m-QAM constellation Ω. It is assumed that the data symbols x_i's are i.i.d with variance σ_x^2. After precoding, precoded symbols are transmitted through n_t transmit antennas over a block-fading channel.

With n_t transmit antennas and n_r receive antennas, the channel is modeled by an $n_t \times n_r$ matrix. For frequency-flat Rayleigh fading, coefficients of the channel matrix are i.i.d. zero-mean circularly symmetric complex Gaussian random variables with variance σ_h^2. The channel is assumed to be block fading with n_c different channel realizations during each codeword. For the kth symbol to be precoded, $\mathbf{x}_k$, the $Nn_t \times Nn_r$ extended channel matrix, $\mathbf{H}_k$, can be written as

$$\mathbf{H}_k = \text{diag}\left\{ \underbrace{\mathbf{H}_k^{[1]}, \ldots, \mathbf{H}_k^{[1]}}_{N/n_s}, \mathbf{H}_k^{[2]}, \ldots, \mathbf{H}_k^{[2]}, \ldots, \mathbf{H}_k^{[n_s]}, \ldots, \mathbf{H}_k^{[n_s]} \right\}, \quad (1)$$

where n_s is the number of distinct channel realizations during N time periods of each codeword. To simplify the notation it is also assumed[a] that n_s divides N. For example, if the length of a codeword is 64 and $n_c = 32$, then choosing $N = 2$ would make $n_s = 1$, whereas choosing $N = 4$ gives $n_s = 2$. Notation $\mathbf{H}_k^{[t]}$ refers to the $n_t \times n_r$ complex matrix k that defines the tth channel realization included in n_s channel realizations. The extended channel input/output relationship is expressed by

$$\mathbf{y}_k = \mathbf{x}_k\mathbf{G}\mathbf{H}_k + \mathbf{w}_k \quad (2)$$

Figure 1 Block diagram of a BICM-MIMO system with a linear precoder and proposed training insertion.

where $\mathbf{y}_k = [\gamma_{(k-1)Nn_r+1}, \gamma_{(k-1)Nn_r+2}, \ldots, \gamma_{(k-1)Nn_r+Nn_r}]$ is the received vector at the kth precoding time period and $\mathbf{w}_k$ is the noise vector with size $1 \times Nn_r$ whose components are i.i.d zero-mean circularly

Figure 2 Spreading a precoded symbol over n_t antennas and N time periods-denoted by $\mathbf{Z}_k = [Z_{1,k}, Z_{2,k}, \ldots, Z_{n_t,k}]^T$.

symmetric Gaussian random variables with variance N_0. It is noted from (2) that although both data and pilot symbols are precoded, the part of the precoder that multiplies the pilot symbols depends on the positions of the pilot symbols in $\mathbf{x}_k$. Equivalently, the design of the pilot symbols is governed by the properties of the precoder used. Since this study adopts the transmission framework and precoder design in [6], it is useful to review the properties of the precoder proposed in [6].

In general, the properties of the precoder in [6] are established by the maximum-likelihood decoding analysis and an assumption of ideal channel interleaving. Specifically, this linear precoder which achieves full diversity order and maximum coding gain satisfies the following two conditions:

• A genie condition, which guarantees orthogonal and equal norm sub-rows in the linear precoding matrix. Each sub-row has size n_t in a precoding matrix with size $Nn_t \times Nn_t$.
• Dispersive nucleo algebraic (DNA) condition, which is based on Proposition 2 in [6], forces null and orthogonal nucleotides with size $s' = N/n_s$. Nucleotides refer to subparts of sub-rows with size s'.
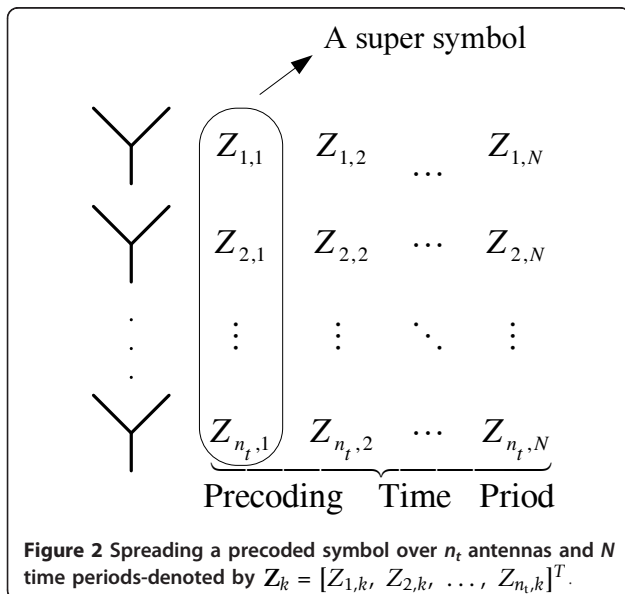
A linear precoder that satisfies the above two sets of conditions is called DNA-cyclo precoder and has the best performance in terms of achieving diversity and coding gains with low complexity receiver when $N \leq n_t$. It is suggested in [6] that to generate one class of such a precoder, a $Ns' \times Ns'$ cyclotomic rotator, denoted by $\boldsymbol{\Phi}$, that satisfies the genie condition is first selected. Then the orthogonal nucleotides are placed inside an $Nn_t \times Nn_t$ matrix and they are separated with null nucleotides. Therefore, the DNA-cyclo precoder matrix can be expressed by subparts of a cyclotomic rotator as follows:

$$\mathbf{G} = \begin{bmatrix} \mathbf{I}_{n_t/s'} \otimes \boldsymbol{\Phi}^{[1][1]} & \cdots & \mathbf{I}_{n_t/s'} \otimes \boldsymbol{\Phi}^{[N][1]} \\ \mathbf{I}_{n_t/s'} \otimes \boldsymbol{\Phi}^{[1][2]} & \cdots & \mathbf{I}_{n_t/s'} \otimes \boldsymbol{\Phi}^{[N][2]} \\ \vdots & \ddots & \vdots \\ \mathbf{I}_{n_t/s'} \otimes \boldsymbol{\Phi}^{[1][Ns']} & \cdots & \mathbf{I}_{n_t/s'} \otimes \boldsymbol{\Phi}^{[N][Ns']} \end{bmatrix} \quad (3)$$

where $\boldsymbol{\Phi}^{[i][j]}$ is the ith sub-row of the jth row of $\boldsymbol{\Phi}$ with size $1 \times s'$, $\mathbf{I}_n$ is an identity matrix with size $n \times n$ and $\otimes$ denotes the Kronecker product.

The properties that shall be useful for the problem considered in this article, which are implied directly from the genie and DNA conditions, are $\boldsymbol{\Phi}\boldsymbol{\Phi}^H = \mathbf{I}_{Ns'}$ and $\boldsymbol{\Phi}^{[i][t]}(\boldsymbol{\Phi}^{[j][t]})^H = \frac{1}{N}\delta(i-j)$. It is also useful to point out that each component of $\boldsymbol{\Phi}$ has an exponential form with a scaling factor of $\frac{1}{\sqrt{Ns'}}$.

The iterative receiver is also shown in Figure 1. The channel estimator produces an estimate of the channel using the minimum MSE (MMSE) criterion based on the training sequence. Details about channel estimation with the proposed method of inserting training sequence shall be given in Section 3. After channel estimation is performed using the training signal, the soft-input soft-output demodulator uses the MMSE criterion to demodulate the data. The soft-output MMSE demodulator computes the extrinsic information for the interleaved bits, $\{\Lambda_{\text{ext}}^{(\tilde{c}_l)}\}$, from the received symbols. To obtain Λ-values, the demodulator exploits the *a priori* information of the coded bits coming from the decoder, $\{\Lambda_{\text{ap}}^{(\tilde{c}_l)}\}$, and the channel estimate $\hat{\mathbf{H}}_k$. In the first iteration, the demodulator assumes that the *a priori* Λ-values are zero, except for the pilot symbols. For the corresponding bits of the pilot symbols, the demodulator uses a large number, say ±100 as their *a priori* Λ-values. The de-interleaved outputs, i.e., $\{\Lambda_{\text{ap}}^{(c_l)}\}$, become the *a priori* Λ-values used in the channel decoder shown in Figure 1 after removing the information of pilot symbols. The channel decoder uses the maximum *a posteriori* probability (MAP) algorithm to compute the extrinsic Λ-values $\{\Lambda_{\text{ext}}^{(c_l)}\}$. for all coded bits, which are used again in the next iteration in the demodulator. In subsequent iterations, soft information from the decoder is used to improve the performance of the channel estimator. The detailed operation of the iterative channel estimator is discussed in the following sections.

3 Training design and channel estimator
As discussed before, the criterion used for training design in this article is the CRB on the channel estimation error. The bound states that the MSE of any unbiased estimator is lower bounded by the trace of inverse of complex Fisher information matrix (FIM) [14]. To derive FIM, the relation between the channel input and channel output during one block-length, i.e., N/n_s time periods, whose corresponding channel matrix is $\mathbf{H}_k^{[t]}$, is of interest. In the following, index k is omitted, since it suffices to consider the transmission of a single precoded symbol for the purpose of channel estimation. With the previously described structure of the precoder, the channel output during one super-symbol time is given by

$$\mathbf{y}^{[i,t]} = \left(\mathbf{I}_{n_r} \otimes \left(\sum_{\tau=1}^{Ns'} \mathbf{x}^{[\tau]} \otimes \boldsymbol{\Phi}^{[i,t][\tau]} \right) \right) \mathbf{h}^{[t]} + \mathbf{w}^{[i,t]}; \quad t = 1, \ldots, n_s, \; i = 1, \ldots, s' \quad (4)$$

where $\mathbf{y}^{[i,t]} = \mathbf{y}^{[(t-1)s'+i]}$ represents the $((t-1)s'+i)$th received symbol during N time periods, with size $n_r \times 1$. Moreover, $\mathbf{h}^{[t]}$ is the column vector formed by vertically stacking the columns of an $n_t \times n_r$ channel realization matrix $\mathbf{H}^{[t]}$ and $\mathbf{x}^{[\tau]}$'s are constructed by splitting $\mathbf{x}$ in Ns' sub-vectors with size $1 \times n_t/s'$. In the following, we call these sub-vectors $\mathbf{x}^{[\tau]}$'s *nucleo* symbols.

It is quite obvious from (4) that, to have all the received super-symbols, $\mathbf{y}^{[i,t]}$, contain training information, there should be at least one pilot nucleo (i.e., n_t/s' pilot symbols) in each group of Ns' nucleos to be precoded.

With the above structure of the proposed training sequence, the number of pilot symbols in Nn_t transmitted symbols would be $N_p = n_p \times n_t/s'$, where n_p nucleo symbols in a symbol to be precoded are assigned to training sequence. Therefore, (4) can be rewritten as

$$\mathbf{y}^{[i,t]} = \left(\mathbf{I}_{n_r} \otimes \left(\sum_{\tau \in \mathcal{I}_d} \mathbf{x}_d^{[\tau]} \otimes \boldsymbol{\Phi}_d^{[i,t][\tau]} + \sum_{\tau \in \mathcal{I}_p} \mathbf{x}_p^{[\tau]} \otimes \boldsymbol{\Phi}_p^{[i,t][\tau]} \right) \right) \mathbf{h}^{[t]} + \mathbf{w}^{[i,t]}, \quad (5)$$

where $\mathcal{I}_d$ and $\mathcal{I}_p$ are sets of indexes from $\{1, \ldots, Ns'\}$, that are assigned to data and pilot nucleos, respectively, and $|\mathcal{I}_d| + |\mathcal{I}_p| = (Ns' - n_p) + n_p = Ns'$. Note that the subscripts "d" and "p" are used to differentiate between data and pilot nucleos. For convenience, the notations $\boldsymbol{\Phi}_p^{[i,t][\tau]}$ and $\boldsymbol{\Phi}_d^{[i,t][\tau]}$ are used to refer to sub-rows of $\boldsymbol{\Phi}$ that are multiplied by pilot and data nucleos, i.e., $\mathbf{x}_p^{[\tau]}$ and $\mathbf{x}_d^{[\tau]}$, respectively. Furthermore, in the

following the notation $\mathbf{T}^{[i,t]}$ is used for $\mathbf{I}_{n_r} \otimes \left(\sum_{\tau \in \mathcal{I}_p} \mathbf{x}_p^{[\tau]} \otimes \mathbf{\Phi}_p^{[i,t]\,[\tau]} \right)$.

The derivation of FIM is given in the next section. Pilot symbols are exploited at the initialization phase and in subsequent iterations considering the special structure of the training sequence. In general, training design can be investigated for these two phases separately. However, for the precoder adopted in this article, the optimal training design obtained for the initialization phase turns out to also be optimal for the iteration phase. Nevertheless, the optimal numbers of pilot nucleos in these two phases of channel estimation are not the same.

3.1 Fisher information matrix

The key steps in deriving the FIM in the *initialization phase* are now given. Without loss of generality we drop superscript t in (5) and perform all the derivations for the first block period (i.e., $t = 1$). Collecting all the observations during the first block period of length s' in a vector ϕ, the FIM for the channel estimation problem at the initialization phase is defined and computed as

$$
\begin{aligned}
\text{FIM}^{\text{init}}(n_p, \mathbf{x}_p, \mathcal{I}_p) &= E_{\varphi,\mathbf{h}}\left\{ \left[\frac{\partial \ln p(\varphi, \mathbf{h})}{\partial \mathbf{h}^*} \right] \left[\frac{\partial \ln p(\varphi, \mathbf{h}))}{\partial \mathbf{h}^*} \right]^H \right\} \\
&= E_{\mathbf{h}}\left\{ E_{\varphi}\left\{ \left[\frac{\partial \ln p(\varphi|\mathbf{h})}{\partial \mathbf{h}^*} \right] \left[\frac{\partial \ln p(\varphi|\mathbf{h})}{\partial \mathbf{h}^*} \right]^H \Big| \mathbf{h} \right\} \right\} \\
&\quad + E_{\mathbf{h}}\left\{ \left[\frac{\partial \ln p(\mathbf{h})}{\partial \mathbf{h}^*} \right] \left[\frac{\partial \ln p(\mathbf{h})}{\partial \mathbf{h}^*} \right]^H \right\}
\end{aligned} \quad (6)
$$

where $\text{FIM}^{\text{init}}(n_p, \mathbf{x}_p, \mathcal{I}_p)$ shows the dependency of FIM on those parameters of interest. Using the i.i.d. assumption on noise and data, $p(\phi|\mathbf{h})$ can be approximated as a complex normal distribution with mean $\mu = [\mu_1^T, \ldots, \mu_{s'}^T]^T$ and covariance $\mathbf{R}_\phi = \text{diag}[\mathbf{R}_1, \ldots, \mathbf{R}_{s'}]$. Moreover, it follows from (5) that $\mu_i = E_\phi\{\mathbf{y}^{[i]}|\mathbf{h}\} = \mathbf{T}^{[i]}\mathbf{h}$ and

$$
\mathbf{R}_i = \mathcal{H}\left(\sigma_x^2 I_{n_t/s'} \otimes \left((\mathbf{\Phi}_d^{[i]})^T (\mathbf{\Phi}_d^{[i]})^* \right) \right) \mathcal{H}^H + N_0 \mathbf{I}_{n_r} \quad (7)
$$

where $\mathcal{H} = (\mathbf{H}^{[1]})^T$ and $\mathbf{\Phi}_d^{[i]}$ is the ith sub-matrix of $\mathbf{\Phi}$ with size $(N s' - n_p) \times s'$ that is assigned to data symbols.

The i.i.d. assumptions on noise and data make the FIM additive. Specifically, $\text{FIM}^{\text{init}}(n_p, \mathbf{x}_p, I_p) = \sum_{i=1}^{s'} \text{FIM}_i^{\text{init}}$. The quantity $\text{FIM}_i^{\text{init}}$ is obtained as follows:

$$
\text{FIM}_i^{\text{init}} = E_{\mathbf{h}}\left\{ E_{\mathbf{y}}\left\{ \frac{\partial \ln p(\mathbf{y}|\mathbf{h})}{\partial \mathbf{h}^*} \left(\frac{\partial \ln p(\mathbf{y}|\mathbf{h})}{\partial \mathbf{h}^*} \right)^H \Big| \mathbf{h} \right\} \right\} + \sigma_h^{-2} \mathbf{I}_{n_t n_r}.
$$

We know that

$$
\ln p(\mathbf{y}|\mathbf{h}) = \text{Constant} - \ln |\mathbf{R}_i| - (\mathbf{y} - \mu_i)^H \mathbf{R}_i^{-1}(\mathbf{y} - \mu_i). \quad (8)
$$

and $\frac{\partial \ln |\mathbf{R}_i|}{\partial h_l^*} = \text{trace}\left(\mathbf{R}_i^{-1} \frac{\partial \mathbf{R}_i}{\partial h_l^*} \right)$. Therefore,

$$
\frac{\partial \mathbf{R}_i}{\partial h_l^*} = \mathcal{H}\left(\sigma_x^2 I_{n_t/s'} \otimes \left((\mathbf{\Phi}_d^{[i]})^T (\mathbf{\Phi}_d^{[i]})^* \right) \right) \Sigma_l^T \quad (9)
$$

where Σ_l is an $n_r \times n_t$ null matrix with only a single element of 1 at position $\left(\left\lfloor \frac{l-1}{n_t} \right\rfloor + 1, (l - 1 \mod n_t) + 1 \right)$. The derivative of the third term in (8) is

$$
\frac{\partial (\mathbf{y} - \mu_i)^H \mathbf{R}_i^{-1}(\mathbf{y} - \mu_i)}{\partial h_l^*} = -\frac{\partial \mu_i^H}{\partial h_l^*} \mathbf{R}_i^{-1}(\mathbf{y} - \mu_i) + (\mathbf{y} - \mu_i)^H \frac{\partial \mathbf{R}_i^{-1}}{\partial h_l^*}(\mathbf{y} - \mu_i)
$$

where $\frac{\partial \mathbf{R}_i^{-1}}{\partial h_l^*} = -\mathbf{R}_i^{-1} \frac{\partial \mathbf{R}_i}{\partial h_l^*} \mathbf{R}_i^{-1}$ and $\frac{\partial \mathbf{R}_i}{\partial h_l^*}$ is given by (9). In addition,

$$
\frac{\partial \mu_i^H}{\partial h_l^*} = \frac{\partial \mathbf{h}^H}{\partial h_l^*}(\mathbf{T}^{[i]})^H = \mathbf{e}_l^T (\mathbf{T}^{[i]})^H
$$

where $\mathbf{e}_l$ is an $n_t n_r \times 1$ null vector with a single element 1 at position l.

Using all the above equations and after some manipulations, one has

$$
\begin{aligned}
(\text{FIM}_i^{\text{init}})_{l,j} &= E_{\mathbf{h}}\{\mathbf{e}_l^T (\mathbf{T}^{[i]})^H \mathbf{R}_i^{-1} \mathbf{T}^{[i]} \mathbf{e}_j \\
&\quad + \text{tr}(\mathbf{R}_i^{-1} \mathcal{H} \mathbf{A}^{[i]} \Sigma_l^T \mathbf{R}_i^{-1} \Sigma_j (\mathbf{A}^{[i]})^H \mathcal{H}^H)\} + \sigma_h^{-2}\delta(l - j),
\end{aligned}
$$

where $\mathbf{A}^{[i]} \equiv \left(\sigma_x^2 I_{n_t/s'} \otimes \left((\mathbf{\Phi}_d^{[i]})^T (\mathbf{\Phi}_d^{[i]})^* \right) \right)$.

Using the fact that $\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{CAB})$ and summing over s' quantities $\text{FIM}_i^{\text{init}}$, the total FIM is given by,

$$
\text{FIM}^{\text{init}}(n_p, \mathbf{x}_p, \mathcal{I}_p) = E_{\mathbf{h}}\left\{ \sum_{i=1}^{s'} \mathbf{R}_i^{-1} \otimes ((\mathbf{X}_p^{[i]})^H \mathbf{X}_p^{[i]}) + \mathbf{R}_i^{-1} \otimes \mathbf{Q}_i \right\} + s' \sigma_h^{-2} \mathbf{I}_{n_t n_r} \quad (10)
$$

where $\mathbf{X}_p^{[i]} = \sum_{\tau \in \mathcal{I}_p} \mathbf{x}_p^{[\tau]} \otimes \mathbf{\Phi}_p^{[i][\tau]}$, and $\mathbf{Q}_i = (\mathbf{A}^{[i]})^T \mathcal{H}^T \mathbf{R}_i^{-1} \mathcal{H}^* (\mathbf{A}^{[i]})^*$.

For designing training sequence, (10) can be simplified further using numerical calculation. Using numerical calculation, it is observed that for a Rayleigh-distributed channel, the matrix $E_{\mathbf{h}}\{\mathbf{R}_i^{-1}\}$ in (10) is approximately a diagonal matrix[b], $\alpha \mathbf{I}_{n_r}$. This observation means that $E_{\mathbf{h}}\{\mathbf{Q}_i\}$ can be approximated by $n_r \alpha \sigma_h^2 (\mathbf{A}^{[i]})^T (\mathbf{A}^{[i]})^*$. Then, by performing the expectation operation and using the factorization property of the Kronecker product, (10) can be represented as

$$E_h \left\{ \sum_{i=1}^{s'} \alpha \mathbf{I}_{n_r} \otimes \left((\mathbf{X}_p^{[i]})^H \mathbf{X}_p^{[i]} \right) + \alpha \mathbf{I}_{n_r} \otimes n_r \alpha \sigma_h^2 (\mathbf{A}^{[i]})^\tau (\mathbf{A}^{[i]})^* \right\} + s' \sigma_h^{-2} \mathbf{I}_{n_r} \otimes \mathbf{I}_{n_t} =$$

$$\mathbf{I}_{n_r} \otimes \left(\sum_{i=1}^{s'} \alpha \left((\mathbf{X}_p^{[i]})^H \mathbf{X}_p^{[i]} \right) + n_r \alpha^2 \sigma_h^2 (\mathbf{A}^{[i]})^\tau (\mathbf{A}^{[i]})^* + s' \sigma_h^{-2} \mathbf{I}_{n_t} \right)$$

Moreover, using the property of the Kronecker product $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$, it follows that $(\mathbf{X}_p^{[i]})^H \mathbf{X}_p^{[i]} = \sum_{\tau \in \mathcal{I}_p} \sum_{\tau' \in \mathcal{I}_p} ((\mathbf{x}_p^{[\tau]})^H \mathbf{x}_p^{[\tau']}) \otimes ((\mathbf{\Phi}_p^{[i][\tau]})^H \mathbf{\Phi}_p^{[i][\tau']})$. Therefore (10) can be further simplified to

$$\text{FIM}^{\text{init}}(n_p, \mathbf{x}_p, \mathcal{I}_p) = \mathbf{I}_{n_r} \otimes$$
$$\left(\alpha \sum_{\tau \in \mathcal{I}_p} \sum_{\tau' \in \mathcal{I}_p} (\mathbf{x}_p^{[\tau]})^H \mathbf{x}_p^{[\tau']} \otimes \sum_{i=1}^{s'} (\mathbf{\Phi}_p^{[i][\tau]})^H \mathbf{\Phi}_p^{[i][\tau']} + n_r \alpha^2 \sigma_h^2 \sum_{i=1}^{s'} (\mathbf{A}^{[i]})^\tau (\mathbf{A}^{[i]})^* + s' \sigma_h^{-2} \mathbf{I}_{n_t} \right) \quad (11)$$

In general, the second term in (11) depends on $\mathcal{I}_p$, but not on the training symbols, whereas the first term depends on both $\mathbf{x}_p$ and $\mathcal{I}_p$. Although both terms depend on n_p, how FIM$^{\text{init}}$ depends on n_p is determined by $\mathcal{I}_p$. Therefore, in the following $\mathcal{I}_p$ and $\mathbf{x}_p$ are first optimized. Then n_p is determined for the optimized $\mathcal{I}_p$.

For the *iteration phase*, specifically the last iteration, estimation and detection are implemented using information about the data symbols as well as the pilot symbols. Thus, the parameter of interest in deriving FIM is $\boldsymbol{\theta} = [\mathbf{h}^T \mathbf{x}_d]^T$. Moreover, $\mu_i = E_\varphi \{\mathbf{y}^{[i]} | \boldsymbol{\theta}\} = \left(\mathbf{I}_{n_r} \otimes \left(\sum_\tau \mathbf{x}^{[\tau]} \otimes \mathbf{\Phi}^{[i][\tau]} \right) \right) \mathbf{h}$ and $\mathbf{R}_i = N_0 \mathbf{I}_{n_r}$. By replacing $\boldsymbol{\theta}$ in (6) for $\mathbf{h}$ and after some manipulations, the FIM for channel estimation in the iteration phase is given by

$$\text{FIM}^{\text{iter}}(n_p, \mathbf{x}_p, \mathcal{I}_p) = N_0^{-1} \mathbf{I}_{n_r} \otimes$$
$$\left(\frac{Ns' - n_p}{N} \sigma_x^2 \mathbf{I}_{n_t} + \sum_{\tau \in \mathcal{I}_p} \sum_{\tau' \in \mathcal{I}_p} (\mathbf{x}_p^{[\tau]})^H \mathbf{x}_p^{[\tau']} \otimes \sum_{i=1}^{s'} (\mathbf{\Phi}_p^{[i][\tau]})^H \mathbf{\Phi}_p^{[i][\tau']} \right) + s' \sigma_h^{-2} \mathbf{I}_{n_r n_t}. \quad (12)$$

3.2 Optimization of training symbols and their positions
This section is first concerned with minimizing the CRB expression for the initialization phase. The minimization is under a constraint on the power budget for the training sequence. Such a constraint is expressed as

$$\sum_{\tau \in \mathcal{I}_p} \left(\mathbf{x}_p^{[\tau]} \otimes \mathbf{\Phi}_p^{[i][\tau]} \right) \left((\mathbf{x}_p^{[\tau]})^H \otimes (\mathbf{\Phi}_p^{[i][\tau]})^H \right) \leq P_t. \quad (13)$$

Using the properties of the precoder employed in this study, the above constraint can be simplified to $\frac{s'}{N} \sum_{\tau' \in \mathcal{I}_p} \mathbf{x}_p^{[\tau]} (\mathbf{x}_p^{[\tau]})^H \leq P_t$. The other obvious constraint is that the training symbols should be selected from QAM constellation Ω. Then, the training symbols, $\mathbf{x}_p$'s and their positions, specified by $\mathcal{I}_p$, are obtained by solving the following constrained optimization problem:

$$\min_{\mathbf{x}_p, \mathcal{I}_p} \text{CRB}^{\text{init}}(n_p, \mathbf{x}_p, \mathcal{I}_p) = \min_{\mathbf{x}_p, \mathcal{I}_p} \text{tr} \left(\text{FIM}^{\text{init}}(n_p, \mathbf{x}_p, \mathcal{I}_p)^{-1} \right)$$
$$\text{s.t.} \begin{cases} \frac{s'}{N} \sum_{\tau \in \mathcal{I}_p} \mathbf{x}_p^{[\tau]} (\mathbf{x}_p^{[\tau]})^H \leq P_t \\ (\mathbf{x}_p^{[\tau]})_j \in \Omega, \ j = 1, \ldots, n_t/s', \ \tau \in \mathcal{I}_p \end{cases} \quad (14)$$

where $(\mathbf{x}_p^{[\tau]})_j$ is the jth pilot symbol in the τth pilot nucleo and the FIM is given in (11).

To proceed, lets consider two separate cases for problem (14): $n_p = 1$ and $n_p \geq 2$. Case 1 ($n_p = 1$): In this case the FIM is simplified to

$$\mathbf{I}_{n_r} \otimes \left(\alpha \left((\mathbf{x}_p^{[\tau]})^H \mathbf{x}_p^{[\tau]} \right) \otimes \sum_{i=1}^{s'} \left((\mathbf{\Phi}_p^{[i][\tau]})^H \mathbf{\Phi}_p^{[i][\tau]} \right) \right.$$
$$\left. + n_r \alpha^2 \sigma_h^2 \sum_{i=1}^{s'} (\mathbf{A}^{[i]})^\tau (\mathbf{A}^{[i]})^* + s' \sigma_h^{-2} \mathbf{I}_{n_t} \right), \quad (15)$$

Because of the shift-invariant property of (15) with respect to τ, τ can be any value in the set $\{1, 2, \ldots, Ns'\}$. For simplicity, set $\tau = 1$ and the superscript τ is omitted. Using the fact that if $\mathbf{X} > 0$ then $\text{tr}(\mathbf{X}^{-1}) \geq \sum_i 1/(\mathbf{X})_{ii}$, the original optimization problem is simplified by minimizing the lower bound of the objective function.

On the other hand, $\sum_{i=1}^{s'} \left((\mathbf{\Phi}_p^{[i][\tau]})^H \mathbf{\Phi}_p^{[i][\tau]} \right) = \frac{1}{N} \mathbf{I}_{s'}$, $\sum_{i=1}^{s'} (\mathbf{A}^{[i]})^T (\mathbf{A}^{[i]})^* = \frac{\sigma_x^4}{s'} \left((\frac{Ns'-1}{N})^2 + (\frac{1}{N})^2 \right) \mathbf{I}_{n_t}$ and the constraint is $\frac{s'}{N} \mathbf{x}_p \mathbf{x}_p^H = \frac{s'}{N} \sum_{j=1}^{n_t/s'} |(\mathbf{x}_p)_j|^2$. Therefore, it is not hard to see that the solution of the simplified optimization problem is $|(\mathbf{x}_p)_1|^2 = |(\mathbf{x}_p)_2|^2 = \cdots = |(\mathbf{x}_p)_{n_t/s'}|^2 = \frac{NP_t}{n_t}$. It means that all pilot symbols should have the same power. For example, one can select corner points of the QAM constellations for the training symbols.

Case 2 ($n_p \geq 2$): In this case there are two options for the placements of pilot nucleos. The first option is to group all pilot nucleos in one single cluster and the second option is to spread pilot nucleos. It can be shown that the CRB is invariant with respect to a shift of the placements of pilot nucleos in both options. Therefore, it suffices to select one cluster or one spread placement. However, the precoder has been designed such that the soft-output demodulator works with uncorrelated inputs and putting pilot nucleos between data nucleos may violate this condition. That condition is satisfied when $\mathbf{A}^{[i]}$ has a diagonal form. The implication of this property is to place pilot nucleos equi-spaced in $\mathbf{x}_k$ and $\mathcal{I}_p = \{i_0 + kn; \ k = 0, \ldots, n_p - 1\}$, where $n = Ns'/n_p$

and $i_0 \in \{1, \ldots, n\}$, which leads to $\mathbf{A}^{[i]} = \sigma_x^2 \frac{Ns'-n_p}{Ns'}\mathbf{I}_{s'}$. In this selection it is supposed that n_p is divisible by Ns'.

Then the FIM in (11) can be represented by

$$\mathbf{I}_{n_r} \otimes \left(\frac{1}{N}\alpha \sum_{\tau \in \mathcal{I}_p}(\mathbf{x}_p^{[\tau]})^H \mathbf{x}_p^{[\tau]} \otimes \mathbf{I}_{s'} + n_r \alpha^2 \sigma_x^4 \sigma_h^2 \frac{1}{s'} \left(\frac{Ns'-n_p}{N}\right)^2 \mathbf{I}_{n_t} + s'\sigma_h^{-2}\mathbf{I}_{n_t} \right) \quad (16)$$

To obtain the above expression of the objective function, the following property has been used:

$$\left(\sum_{i=1}^{s'}(\boldsymbol{\Phi}_p^{[i][\tau]})^H \boldsymbol{\Phi}_p^{[i][\tau']} \right)_{n,l} = \begin{cases} \frac{1}{N}, & \tau = \tau'; \; n = l \\ 0, & \text{otherwise} \end{cases} \quad (17)$$

Moreover, the only term that depends on the training symbols is $\sum_{\tau \in \mathcal{I}_p}(\mathbf{x}_p^{[\tau]})^H \mathbf{x}_p^{[\tau]}$ in (16). Finally, using the constraint on training power, which can be written as

$$\frac{s'}{N} \sum_{\tau \in \mathcal{I}_p} \sum_{j=1}^{n_t/s'} |(\mathbf{x}_p^{[\tau]})_j|^2 \leq P_t, \quad (18)$$

the solution is given by $\sum_{\tau} |(\mathbf{x}_p^{[\tau]})_j|^2 = \frac{NP_t}{n_t}; \; j = 1, \ldots, n_t/s'$.

Now consider the training design for the iteration phase. Observe that all the terms in (12) have diagonal forms with equal diagonal elements, except $\sum_{\tau \in \mathcal{I}_p} \sum_{\tau' \in \mathcal{I}_p}(\mathbf{x}_p^{[\tau]})^H \mathbf{x}_p^{[\tau']} \otimes \sum_{i=1}^{s'}(\boldsymbol{\Phi}_p^{[i][\tau]})^H \boldsymbol{\Phi}_p^{[i][\tau']}$. This means that the solution of problem (14), but with $\text{FIM}^{\text{init}}(n_p, \mathbf{x}_p, \mathcal{I}_p)$ replaced by $\text{FIM}^{\text{iter}}(n_p, \mathbf{x}_p, \mathcal{I}_p)$, is to choose equal diagonal elements for this term. Therefore, the training sequence designed for the initialization is also optimal for the iteration phase.

In summary, by selecting pilot nucleos such that the sum of the powers of their corresponding pilot symbols with the same indexes are equal, the bound on CRB is minimized. The above condition can give different selections for pilot symbols from a two-dimensional constellation. It should be pointed out, however, that not all selections guarantee that pilot symbols belong to standard QAM constellations.

3.3 Determination of the number of the training symbols

For block-fading channels, the number of pilot nucleos, i.e., n_p, should be as small as possible that meets the power constraint. Using a larger value for n_p wastes bandwidth and does not change the system performance.

The optimum numbers of the training symbols in the initialization phase and iteration phase are not the same. This is explained as follows. At the initialization, by looking at (7), it is observed that the first term in (11) is

an increasing function of n_p. However, the second term is a decreasing function of n_p that is multiplied by n_r. Therefore, n_p that minimizes the CRB are determined by the summation of these two terms, which is also determined by the value of n_r. Table 1 gives several examples of optimal n_p for different sets of n_t, n_r and N. For the iteration phase, the expression in (12) means that the CRB in the iteration phase always increases by increasing n_p. Since it is assumed that there is perfect information about the data symbols in the iteration phase, which is not the case in reality, it is most appropriate to select n_p considering only the initialization phase.

To demonstrate the optimal training design, Figure 3 shows a graphical structure for a simple example, where $P_t = 4\sigma_x^2, n_p = 2, N = 2, n_t = 4$ and $n_r = 2$. In this example, $n_s = 1$. Then the size of pilot nucleos should be $n_t/s' = 2$, where $s' = N/n_s = 2$.

3.4 Channel estimation

For the channel estimation task, one can view the received vector during one block length as $\boldsymbol{\varphi}^{[t]} = [(\mathbf{y}^{[1,t]})^T, (\mathbf{y}^{[2,t]})^T, \ldots, (\mathbf{y}^{[s',t]})^T]^T$.

At the initialization, the mean and covariance matrix of this vector are given in Section 3.1. By treating the data symbols as nuisance parameters, the MMSE channel estimate can be found as [14]

$$\hat{\mathbf{h}}^{[t]} = \sigma_h^2 \mathbf{T}^H(\sigma_h^2 \mathbf{T}\mathbf{T}^H + \mathbf{R}_{\varphi^{[t]}})\boldsymbol{\varphi}^{[t]} \quad (19)$$

where $\mathbf{T} = [(\mathbf{T}^1)^T, \ldots, (\mathbf{T}^{[s']})^T]^T$.

In the 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. Therefore, the interleaved $\{\Lambda_{\text{ext}}^{(c_l)}\}$ from the decoder are fed back to the estimator to calculate the expected values of the data symbols, i.e., $E\{\mathbf{x}_d\}$. The entries of $E\{\mathbf{x}_d\}$ are calculated using $\{\Lambda_{\text{ap}}^{(\tilde{c}_l)}\}$ at each iteration by $E\{(\mathbf{x}_d)_i\} = \sum_{x \in \Omega} x \cdot p((\mathbf{x}_d)_i = x)$. The detailed derivations of the probability $p((\mathbf{x}_d)_i = x)$ from Λ-values are given in

Table 1 Optimum n_p for several sets of parameters $\{n_t, n_r, N\}$

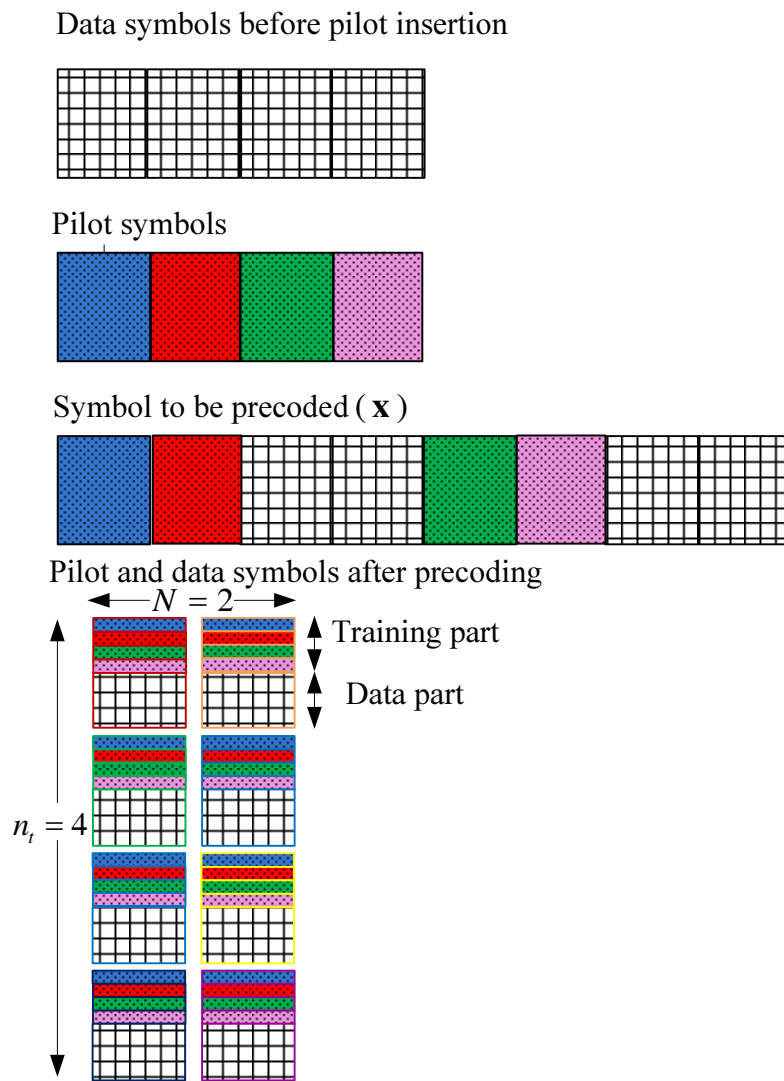| n_t	n_r	N	n_p
2	2	2	1
4	2	2	2
4	2	4	4
4	4	2	1
4	4	4	1

Figure 3 Structure of the proposed scheme for the training sequence-when $N = 2$, $n_t = 4$, $n_r = 2$ and $n_p = 2$

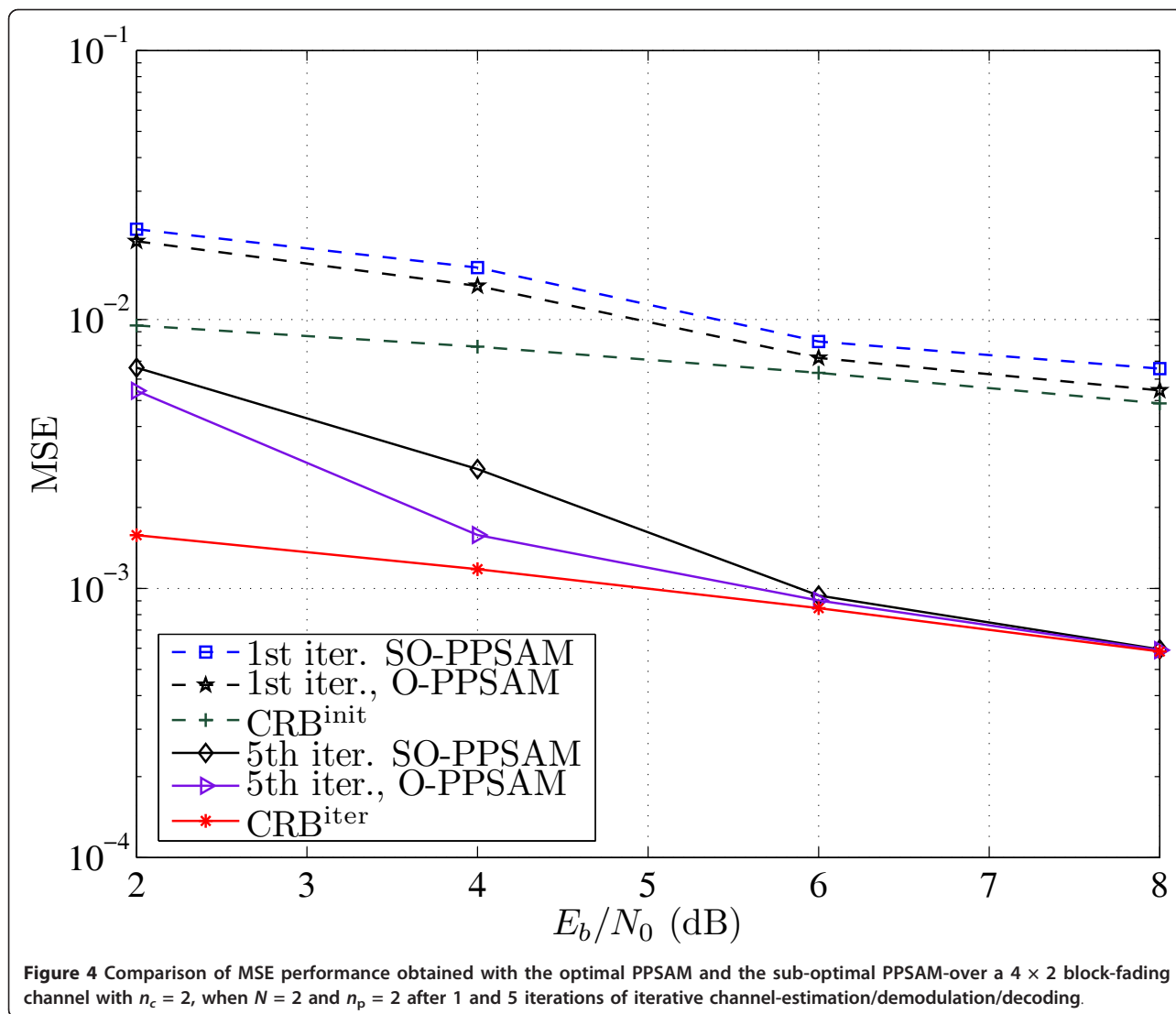[15] (note that the calculation depends on the mapping rule in Ω).

To verify the results obtained in this section, Section 4 compares numerically the MSE performance of the above channel estimator obtained with the optimal and suboptimal training sequences.

4 Illustrative results

In this section, the frame-error-rate (FER) and MSE performances of BICM-MIMO systems using a MMSE iterative channel estimator are presented. The space-time precoder is the DNA-cyclo precoder that satisfies the properties outlined in Section 2. We consider quadrature phase-shift keying (QPSK) modulation with Gray mapping.

The MSE performance of a BICM-MIMO for a codeword length of 4×1024 bits is shown for a 4×2 block-fading MIMO channel in Figure 4, when $n_c = 2$. In this figure, E_b is the energy per information bit. The code used is the 16-state convolutional code with generator polynomials (23, 35) in octal form. In Figure 4, the MSE curves are obtained after 1 and 5 iterations of the iterative channel estimation/demodulation/decoding, with the following cyclotomic rotator [16]:
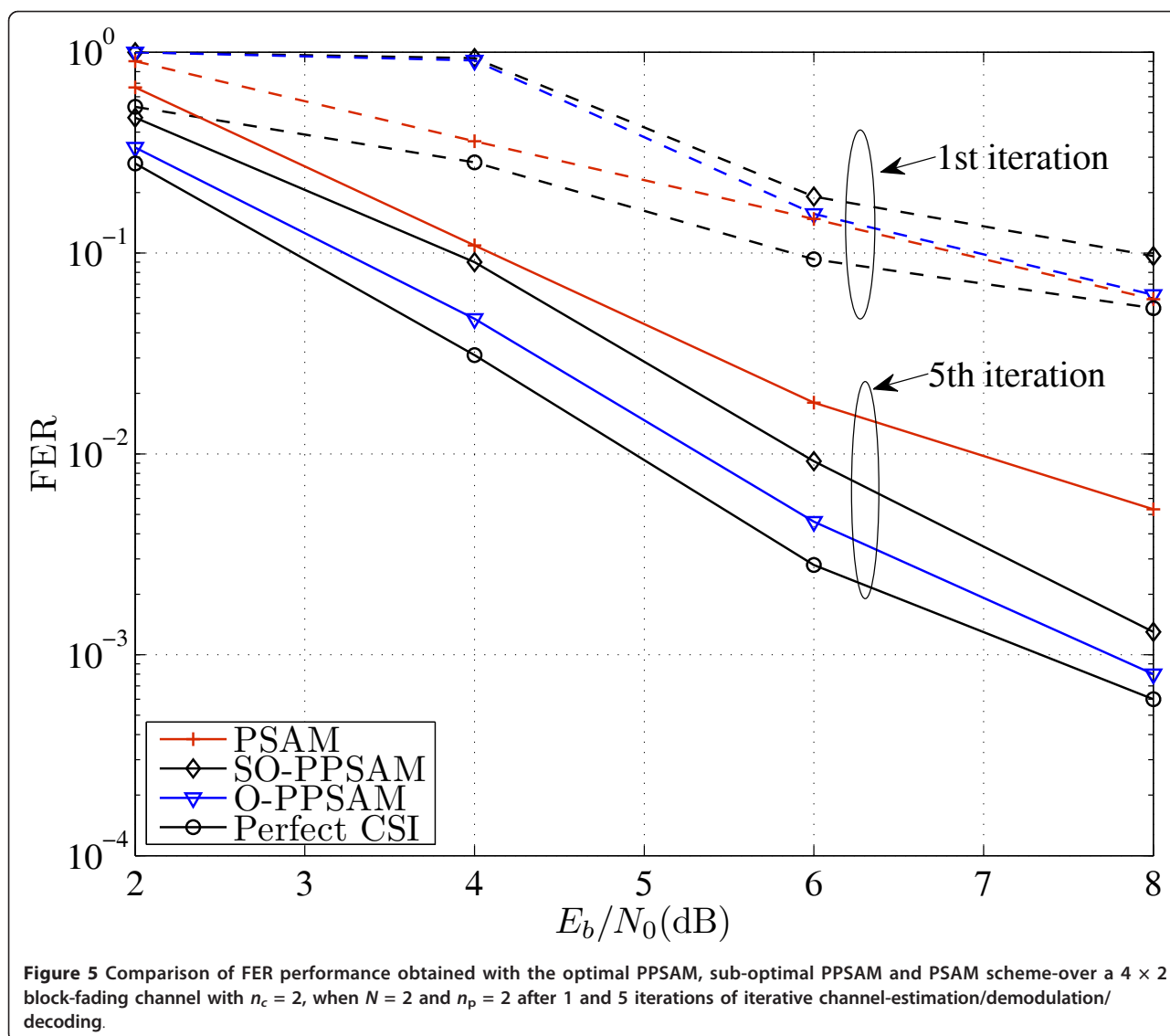
$$\Phi = \frac{1}{2} \begin{bmatrix} 1 & 1 & e^{j6\pi/15} & -e^{j6\pi/15} \\ e^{j2\pi/15} & je^{j2\pi/15} & -e^{j8\pi/15} & je^{j8\pi/15} \\ e^{j4\pi/15} & -e^{j4\pi/15} & e^{j10\pi/15} & e^{j10\pi/15} \\ e^{j6\pi/15} & -je^{j6\pi/15} & -e^{j12\pi/15} & -je^{j12\pi/15} \end{bmatrix}$$

Figure 4 Comparison of MSE performance obtained with the optimal PPSAM and the sub-optimal PPSAM-over a 4 × 2 block-fading channel with $n_c = 2$, when $N = 2$ and $n_p = 2$ after 1 and 5 iterations of iterative channel-estimation/demodulation/decoding.

and when the setting for N, n_s, n_p and P_t in Figure 3 are used. The channel is generated randomly and is assumed to be Rayleigh distributed. For the purpose of comparison, the results for MSE performances of the optimal PPSAM, denoted by O-PPSAM and the suboptimal PPSAM, denoted by SO-PPSAM as well as the CRB are shown in Figure 4. For SO-PPSAM, two pilot nucleos are inserted as one cluster in front of data nucleos in a symbol to be precoded. In contrast, in the case of O-PPSAM, the optimized training sequence embeds the pilot nucleos at the first and third positions of $Ns' = 4$ positions for nucleos. The MSE curves show that the performance of the optimal scheme is better than the sub-optimum scheme for the first iteration (i.e., initialization). In fact the MSE performance of the proposed scheme closely approaches the CRB at high E_b/N_0 after 5 iterations.

In Figure 5, the FER performance of the system with the PPSAM schemes is compared with the conventional PSAM training scheme for the same system parameters as in Figure 4. The top curve is the FER performance of the system with the conventional PSAM training scheme. Note that for a fair comparison, the training scheme in PSAM also meets the training power constraint as trace $(\mathbf{X}_p \mathbf{X}_p^H) = P_t$, where $\mathbf{X}_p$ is the training matrix placed at the beginning of each block of the precoded symbols. The optimal option for PSAM scheme in terms of minimizing the FER as proposed in [11] is to select $\mathbf{X}_p$ to have orthogonal columns. The simplest option is $\sqrt{2 \times \sigma_x^2/n_t}\mathbf{I}_{n_t} = \sqrt{\sigma_x^2}\mathbf{I}_{n_t}$, which results in the same power budget as that of the proposed scheme.

As can be seen from Figure 5, the O-PPSAM scheme offers 0.5 dB performance gain as compared to the SO-

Figure 5 Comparison of FER performance obtained with the optimal PPSAM, sub-optimal PPSAM and PSAM scheme-over a 4 × 2 block-fading channel with n_c = 2, when N = 2 and n_p = 2 after 1 and 5 iterations of iterative channel-estimation/demodulation/decoding.

PPSAM scheme at FER = 10^{-2}. In comparison with PSAM, the performance of the PSAM scheme is about 0.5-1.5 dB worse than the proposed scheme depending on E_b/N_0 after 5 iterations. This is expected because the pilot information is embedded in the precoded symbols for the proposed scheme and not for the PSAM scheme. In this way, the demodulator can also make use of this information. Note, however, that for the first iteration, since there is no information about data, PSAM works the best. More importantly, while the proposed scheme uses a little bandwidth for training information (for the system considered in this figure the training overhead is $n_p \times n_t/s' = 4$), the training overhead of PSAM scheme is $n_t \times n_t = 16$, which is quadruple. To investigate the effect of the number of transmit antennas, two different systems, one with 2 × 2 channel and one with 4 × 2 MIMO channel, are compared in Figures 6 and 7 in

terms of MSE and FER, respectively. For both channels, n_p = 2 and the optimum scheme are used when N = 2, while other system parameters are the same as those used for Figure 4. As can be seen from Figure 6, the MSE of the channel estimation increases when increasing the number of transmit antennas. This is expected because there are more channels to be estimated for the same amount of training information and power as done in the comparison. Nevertheless, the gain in diversity by using more antennas can still improve the overall FER performance as seen in Figure 7.

5 Conclusion

In this article, a new training design for a BICM-MIMO system over a block-fading channel has been proposed. The design inserts pilot symbols into the data symbols before precoding. The new training sequence improves
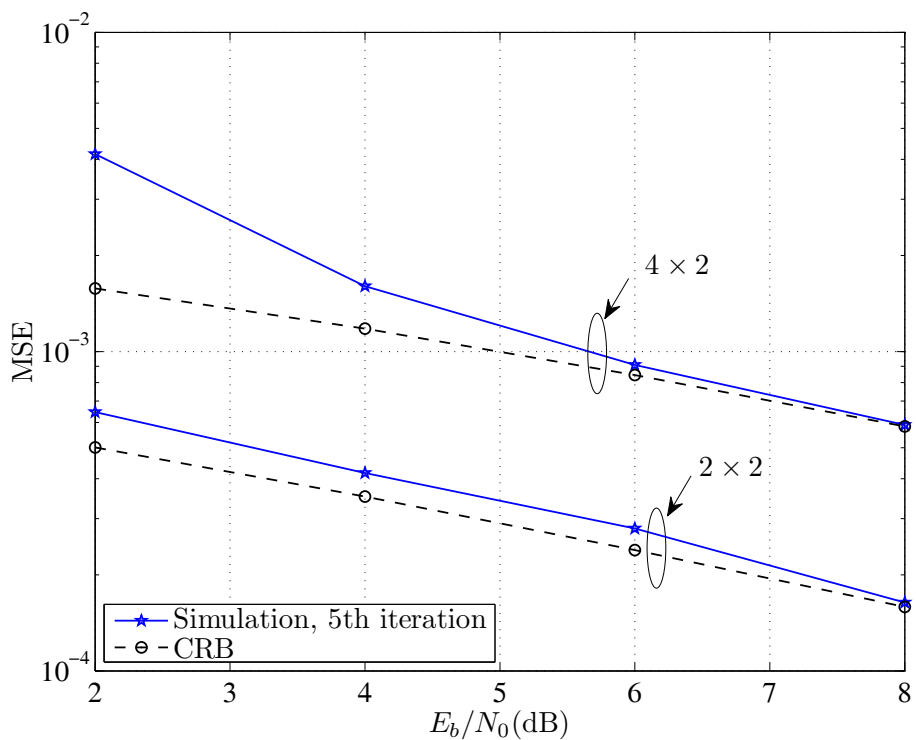
Figure 6 Comparison of MSE performance obtained with the optimal PPSAM for 2 × 2 and 4 × 2 block-fading channels with n_c = 2, when N = 2 and n_p = 2 after 5 iterations of iterative channel-estimation/demodulation/decoding.
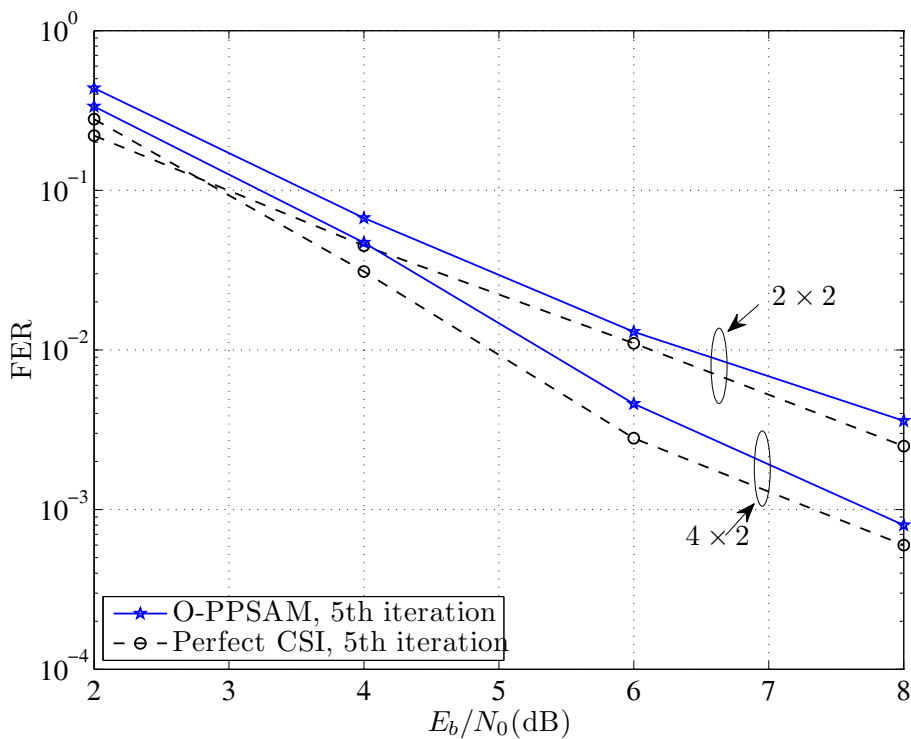

Figure 7 Comparison of FER performance obtained with the optimal PPSAM for 2 × 2 and 4 × 2 block-fading channels with n_c = 2, when N = 2 and n_p = 2 after 5 iterations of iterative channel-estimation/demodulation/decoding.

bandwidth efficiency as compared to the conventional PSAM scheme and can also be used by the demodulator in the receiver. In order to design the optimal training symbols and their positions, the CRB on the channel estimations at the initialization and at the iteration phases are minimized. Compared to PSAM, performance improvement achieved with the proposed training is about 1.5 dB at a FER level of 10^{-2}.

Endnotes

[a]In practice, since n_s is typically an approximated value over some range and since N can be selected, such an assumption can be fulfilled. [b]Using the matrix inversion lemma, one has $\mathbf{R}_i^{-1} = (\mathcal{H}\mathbf{A}^{[i]}\mathcal{H}^H + N_0\mathbf{I}_{n_r})^{-1} = N_0^{-1}\mathbf{I}_{n_r} + N_0^{-2}\mathcal{H}\mathbf{A}^{[i]}\mathcal{H}^H(I_{n_r} + N_0^{-1}\mathcal{H}\mathbf{A}^{[i]}\mathcal{H}^H)^{-1}$. Therefore, for high SNR, $E\{\mathbf{R}_i^{-1}\}$ can be approximated by $N_0^{-1}\mathbf{I}_{n_r}$.

Author details
[1]TRLabs, Saskatoon, Canada [2]Department of Electrical and Computer Engineering, University of Saskatchewan 57 Campus Drive, Saskatoon, SK S7N5A9, Canada

References
1. G Caire, S Shamai, On the achievable throughput of a multiantenna Gaussian broadcast channel. IEEE Trans Inf Theory. **49**(7), 1691–1706 (2003). doi:10.1109/TIT.2003.813523
2. SM Alamouti, A simple transmit diversity technique for wireless communications. IEEE J Sel Areas Commun. **16**(8), 1451–1458 (1998). doi:10.1109/49.730453
3. V Tarokh, N Seshadri, AR Calderbank, Space-time codes for high data rate wireless communication: performance criterion and code construction. IEEE Trans Inf Theory. **44**(2), 744–765 (1998). doi:10.1109/18.661517
4. J Boutros, E Viterbo, Signal space diversity: a power and bandwidth eficient diversity technique for the Rayleigh fading channel. IEEE Trans Inf Theory. **44**(4), 1453–1467 (1998). doi:10.1109/18.681321
5. J Boutros, N Gresset, L Brunel, Turbo coding and decoding for multiple antenna channels, in *International Symposium on Turbo Codes and Related Topics*, Brest, France, pp. 1–8 (2003)
6. N Gresset, L Brunel, J Boutros, Space-time coding techniques with bit-interleaved coded modulations for MIMO block-fading channels. IEEE Trans Inf Theory. **54**(5), 2156–2178 (2008)
7. N Gresset, JJ Boutros, L Brunel, Optimal linear precoding for BICM over MIMO channels, ISIT, 66, Chicago, IL, (September 2004)
8. M Coldrey, P Bohlin, Training-based MIMO systems, Part I: performance comparison. IEEE Trans Signal Process. **55**(11), 5464–5476 (2007)
9. M Nicoli, S Ferrara, U Spagnolini, Soft-iterative channel estimation: methods and performance analysis. IEEE Trans Signal Process. **55**(6), 2993–3006 (2007)
10. M Dong, L Tong, BM Sadler, Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels. IEEE Trans Signal Process. **52**(5), 1403–1418 (2004). doi:10.1109/TSP.2004.826182
11. G Taricco, E Biglieri, Space-time decoding with imperfect channel estimation. IEEE Trans Wirel Commun. **4**(4), 1874–1888 (2005)
12. Y Huang, JA Ritcey, Joint iterative channel estimation and decoding for bit-interleaved coded modulation over correlated fading channels. IEEE Trans Wirel Commun. **4**(5), 2549–2558 (2005)
13. P Piantanida, SM Sadough, On the outage capacity of a practical decoder accounting for channel estimation inaccuracies. IEEE Trans Commun. **57**(5), 1341–1350 (2009)
14. SM Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, (Prentice-Hall PTR, New Jersey, 1993)
15. MA Khalighi, JJ Boutros, Semi-blind channel estimation using the EM algorithm in iterative MIMO APP detectors. IEEE Trans Wirel Commun. **5**(11), 3165–3173 (2006)
16. GM Kraidy, P Rossi, Full-diversity iterative MMSE receivers with space-time precoders over block-fading MIMO channels, in *Proc IEEE Int Conf Wireless Commun and Signal Processing*, Suzhou, pp. 1–5 (2010)