A Joint Design of Code and Training Sequence for Frequency-Selective Block Fading Channels with Partial CSI

Po-Ning Chen
Dept. of Electrical Engineering
National Chiao-Tung University
Taiwan, ROC
Email: poning@faculty.nctu.edu.tw

Chia-Lung Wu
Hardware Development Division
Mobile Devices Inc.
Taiwan, ROC
Email: clwu@banyan.cm.nctu.edu.tw

Mikael Skoglund
School of Electrical Engineering
Royal Institute of Technology
Stockholm, Sweden
Email: skoglund@ee.kth.se

***Abstract*—In this paper, we propose an iterative algorithm to jointly design codes and training sequences for frequency-selective block fading channels with partial channel state information (CSI) at the receiver. After showing that the maximum-likelihood (ML) decoding metric over channels with partial CSI can be well approximated by the joint maximum-likelihood (JML) decoding metric for combined channel estimation and data detection, we propose to use the JML criterion to search for good codes and training sequences in an iterative fashion. Simulations show that the code and training sequence found by our method can outperform a typical system using a channel code with a separately designed training sequence, in particular when codes of low rates are considered.**

***Keywords*—Fading channels, partial channel state information, joint maximum-likelihood, wireless communication**

I. Introduction

In general, the receiver structures can be classified into two categories: coherent receivers and non-coherent receivers. For a coherent receiver, channel estimation is performed first, and then the channel decoder uses the estimated channel information as the true one when performing the decoding. For a non-coherent receiver, no channel estimation device is needed as the decoding is performed without channel state information; hence, no training sequence is necessary for a non-coherent receiver.

The design of a non-coherent system [1–4] usually assumes perfect synchronization, i.e., the receiver can find the exact codeword margins. However, since frame synchronization and channel estimation are often done based on the same training sequence or pilot signals, such an assumption may not be justifiable. As an example, every 148-bit normal burst of GSM signals includes a 26-bit training sequence. Without the information of the codeword margin, joint decoding of the code and training sequence in this 148-bit burst may be technically infeasible. Thus, for a non-coherent system without a training sequence, the assumption of perfect synchronization may be lack of footing.

For this reason, we consider in this work a system with a training sequence for frame-synchronization and channel estimation. Since the estimated channel state information (CSI) is never perfect because of the additive noise, the receiver is termed a *partial coherent* receiver. The first paper about a partial coherent receiver is perhaps [6], which did not draw much attention for long. Recently, several works [7–11] have re-examined this idea. In particular, the authors in [9–11] consider the imperfection of the estimated CSI at receivers, searching for good constellations for certain modulations. Different from these works, we proposed to jointly design the code and training sequence by maximizing the system performance through an iterative search algorithm.

Throughout the paper, the following notations will be used: For a matrix $\mathbb{X}$, $\det|\mathbb{X}|$ is its determinant; $\mathrm{tr}(\mathbb{X})$ is its trace; $\mathbb{X}^{\mathrm{H}}$ denotes its Hermitian transpose.

II. System Model

In our system setting, a signal $\boldsymbol{b} = [b_1, \ldots, b_N]^{\mathrm{T}}$ is transmitted over a frequency-selective block fading (specifically, quasi-static fading) channel of memory order $P-1$. Here, the superscript "T" represents the vector (or matrix) transpose operation. For $1 \le i \le N$, we restrict that b_i is the output of constant-amplitude 2^M-PSK modulation, i.e., $|b_i|^2 = 1$, where $M > 0$. Among the N components in $\boldsymbol{b}$, the first T components are the training sequence and hence are known to both the transmitter and receiver, while the latter $N-T$ ones are used to transmit the data. By denoting $\mathbb{B}$ as

$$\mathbb{B} = \begin{bmatrix} \mathbb{B}_P \\ \mathbb{B}_D \end{bmatrix},$$

which is formed by a $(T\times P)$ matrix $\mathbb{B}_P$ and a $((L-T)\times P)$ matrix $\mathbb{B}_D$ with $L = N+P-1$, where

$$\mathbb{B}_P \triangleq \begin{bmatrix} b_1 & \cdots & 0 \\ b_2 & \ddots & \vdots \\ b_3 & \ddots & 0 \\ b_4 & \ddots & b_1 \\ \vdots & \ddots & \vdots \\ b_T & \cdots & b_{T-P+1} \end{bmatrix}, \mathbb{B}_D \triangleq \begin{bmatrix} b_{T+1} & \cdots & b_{T-P+2} \\ \vdots & \ddots & b_{T-P+3} \\ b_N & \ddots & \vdots \\ 0 & \ddots & b_T \\ \vdots & \ddots & \vdots \\ 0 & \cdots & b_N \end{bmatrix},$$

the received signal $\boldsymbol{y}$ can be formulated by

$$\boldsymbol{y} = \mathbb{B}\boldsymbol{h} + \boldsymbol{n} \tag{1}$$

where $\boldsymbol{n}$ is zero-mean circular symmetric complex Gaussian distributed with correlation matrix $\sigma_n^2\mathbb{I}$, and $\boldsymbol{h} = [h_1, \ldots, h_P]^T$ is the channel taps that remain constant during a L-symbol transmission block (and may change across blocks). Throughout the paper, $\mathbb{I}$ will be used to denote the identity matrix of a proper size. It is assumed that perfect frame synchronization can be achieved, and adequate guard periods are added between consecutive transmission blocks so that there is no inter-block interference. Notably, both the transmitter and receiver know nothing about the channel coefficients $\boldsymbol{h}$ except the multipath parameter P.

A. On the arrangement of training sequence

Generally speaking, the training sequence does not have to be placed at the beginning of $\boldsymbol{b}$ but can be distributed over the entire transmission signal. However, for channels suffering additive white Gaussian noise with unknown constant $\boldsymbol{h}$, such a placement may be justified by the following argument.

Based on the system model in (1), we can divide the received signal $\boldsymbol{y}$ into two parts:

$$\boldsymbol{y}_P = \mathbb{B}_P\boldsymbol{h} + \boldsymbol{n}_P$$

and

$$\boldsymbol{y}_D = \mathbb{B}_D\boldsymbol{h} + \boldsymbol{n}_D,$$

where $\boldsymbol{n}^{\mathtt{H}} = \begin{bmatrix}\boldsymbol{n}_P^{\mathtt{H}} & \boldsymbol{n}_D^{\mathtt{H}}\end{bmatrix}$ and superscript "$\mathtt{H}$" denotes the Hermitian transpose operation. Then the least square estimate of $\boldsymbol{h}$ given $\mathbb{B}_P$ and $\boldsymbol{y}_P$ is

$$\hat{\boldsymbol{h}} = (\mathbb{B}_P^{\mathtt{H}}\mathbb{B}_P)^{-1}\mathbb{B}_P^{\mathtt{H}}\boldsymbol{y}_P.$$

Here, we implicitly assume that $T \geq P$. Denote by $\tilde{\boldsymbol{h}} = \boldsymbol{h} - \hat{\boldsymbol{h}}$ the estimation error. We then derive

$$\begin{aligned}\tilde{\boldsymbol{h}} = \boldsymbol{h} - \hat{\boldsymbol{h}} &= \boldsymbol{h} - (\mathbb{B}_P^{\mathtt{H}}\mathbb{B}_P)^{-1}\mathbb{B}_P^{\mathtt{H}}\boldsymbol{y}_P \\ &= \boldsymbol{h} - (\mathbb{B}_P^{\mathtt{H}}\mathbb{B}_P)^{-1}\mathbb{B}_P^{\mathtt{H}}(\mathbb{B}_P\boldsymbol{h} + \boldsymbol{n}_P) \\ &= -(\mathbb{B}_P^{\mathtt{H}}\mathbb{B}_P)^{-1}\mathbb{B}_P^{\mathtt{H}}\boldsymbol{n}_P,\end{aligned}$$

which implies $E[-(\mathbb{B}_P^{\mathtt{H}}\mathbb{B}_P)^{-1}\mathbb{B}_P^{\mathtt{H}}\boldsymbol{n}_P] = \boldsymbol{0}$ and

$$\begin{aligned}\mathbb{C}_{\tilde{\boldsymbol{h}}} = E[\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}^{\mathtt{H}}] &= E[(\mathbb{B}_P^{\mathtt{H}}\mathbb{B}_P)^{-1}\mathbb{B}_P^{\mathtt{H}}\boldsymbol{n}_P\boldsymbol{n}_P^{\mathtt{H}}\mathbb{B}_P(\mathbb{B}_P^{\mathtt{H}}\mathbb{B}_P)^{-1}] \\ &= \sigma_n^2(\mathbb{B}_P^{\mathtt{H}}\mathbb{B}_P)^{-1}.\end{aligned}$$

Thus, $\tilde{\boldsymbol{h}}$ is zero-mean circular symmetric Gaussian distributed with covariance matrix $\sigma_n^2(\mathbb{B}_P^{\mathtt{H}}\mathbb{B}_P)^{-1}$. We can then obtain a well-known lower bound of mean square error $E[\|\tilde{\boldsymbol{h}}\|^2]$ [2], [5] as

$$E[\|\tilde{\boldsymbol{h}}\|^2] = \operatorname{tr}\left(E[\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}^{\mathtt{H}}]\right) = \sigma_n^2\operatorname{tr}\left((\mathbb{B}_P^{\mathtt{H}}\mathbb{B}_P)^{-1}\right) \geq \frac{\sigma_n^2 P}{T}, \tag{2}$$

with equality holding when $\mathbb{B}_P^{\mathtt{H}}\mathbb{B}_P = T\mathbb{I}$. This then shows that to place the training sequence at the beginning of $\boldsymbol{b}$, together with $\mathbb{B}_P^{\mathtt{H}}\mathbb{B}_P = T\mathbb{I}$, can yield the minimum $E[\|\tilde{\boldsymbol{h}}\|^2]$. In the simulation section, all the training sequences are chosen to satisfy $\mathbb{B}_P^{\mathtt{H}}\mathbb{B}_P = T\mathbb{I}$ so that the minimum $E[\|\tilde{\boldsymbol{h}}\|^2]$ can be achieved.

III. Decoding Criterion for a Receiver with Partial CSI

A. ML decoding criterion

From the channel model in (1), we obtain

$$\boldsymbol{y}_D = \mathbb{B}_D\hat{\boldsymbol{h}} + \mathbb{B}_D(\boldsymbol{h} - \hat{\boldsymbol{h}}) + \boldsymbol{n}_D. \tag{3}$$

Hence, given $\mathbb{B}_D$ and $\hat{\boldsymbol{h}}$, vector $\boldsymbol{y}_D$ is complex Gaussian distributed with mean $E[\boldsymbol{y}_D] = \mathbb{B}_D\hat{\boldsymbol{h}}$ and covariance matrix

$$\mathbb{C} = \sigma_n^2\mathbb{I} + \mathbb{B}_D\mathbb{C}_{\tilde{\boldsymbol{h}}}\mathbb{B}_D^{\mathtt{H}} = \sigma_n^2\left(\mathbb{I} + \mathbb{B}_D(\mathbb{B}_P^{\mathtt{H}}\mathbb{B}_P)^{-1}\mathbb{B}_D^{\mathtt{H}}\right).$$

By Sylvester's determinant theorem, we have

$$\det(\mathbb{C}) = \sigma_n^{2P}\det\left(\mathbb{I} + (\mathbb{B}_P^{\mathtt{H}}\mathbb{B}_P)^{-1}\mathbb{B}_D^{\mathtt{H}}\mathbb{B}_D\right).$$

Together with

$$\mathbb{C}^{-1} = \sigma_n^{-2}\left(\mathbb{I} - \mathbb{B}_D(\mathbb{B}_D^{\mathtt{H}}\mathbb{B}_D + \mathbb{B}_P^{\mathtt{H}}\mathbb{B}_P)^{-1}\mathbb{B}_D^{\mathtt{H}}\right),$$

the ML decoding criterion for the receiver with partial CSI should be

$$\begin{aligned}\hat{\boldsymbol{b}}_{\mathrm{ML}} &= \arg\max_{\mathbb{B}_D}\Pr(\boldsymbol{y}_D|\mathbb{B}_D, \hat{\boldsymbol{h}}) \\ &= \arg\max_{\mathbb{B}_D}\left(\frac{\exp\{-(\boldsymbol{y}_D - \mathbb{B}_D\hat{\boldsymbol{h}})^{\mathtt{H}}\mathbb{C}^{-1}(\boldsymbol{y}_D - \mathbb{B}_D\hat{\boldsymbol{h}})\}}{\pi^P\det(\mathbb{C})}\right) \\ &= \arg\min_{\mathbb{B}_D}\Big(\|\boldsymbol{y}_D - \mathbb{B}_D\hat{\boldsymbol{h}}\|^2 \\ &\qquad -(\boldsymbol{y}_D - \mathbb{B}_D\hat{\boldsymbol{h}})^{\mathtt{H}}\mathbb{Q}_B(\boldsymbol{y}_D - \mathbb{B}_D\hat{\boldsymbol{h}}) \\ &\qquad +\sigma_n^2\log\det\left(\mathbb{I} + (\mathbb{B}_P^{\mathtt{H}}\mathbb{B}_P)^{-1}\mathbb{B}_D^{\mathtt{H}}\mathbb{B}_D\right)\Big),\end{aligned} \tag{4}$$

where

$$\mathbb{Q}_B = \mathbb{B}_D\left(\mathbb{B}_D^{\mathtt{H}}\mathbb{B}_D + \mathbb{B}_P^{\mathtt{H}}\mathbb{B}_P\right)^{-1}\mathbb{B}_D^{\mathtt{H}} = \mathbb{B}_D\left(\mathbb{B}^{\mathtt{H}}\mathbb{B}\right)^{-1}\mathbb{B}_D^{\mathtt{H}}.$$

We then examine the ML decoding criterion in (4) in two extreme cases: no CSI (by which we mean $\mathbb{B}_P$ is a $T \times P$ all-zero matrix) and perfect CSI (by which we mean $\tilde{\boldsymbol{h}} = \boldsymbol{0}$ with probability one).

When $\mathbb{B}_P$ is a $T \times P$ all-zero matrix, (4) can be reduced to the well-known GLRT criterion, i.e.,

$$\begin{aligned}&\hat{\boldsymbol{b}}_{\text{ML, no CSI}} \\ &= \arg\min_{\mathbb{B}_D}\left\{-\boldsymbol{y}_D^{\mathtt{H}}\mathbb{B}_D\left(\mathbb{B}_D^{\mathtt{H}}\mathbb{B}_D\right)^{-1}\mathbb{B}_D^{\mathtt{H}}\boldsymbol{y}_D - \sigma_n^2\log\det(\mathbb{C}^{-1})\right\} \\ &= \arg\min_{\mathbb{B}_D}\left\|\mathbb{P}_D^{\perp}\boldsymbol{y}_D\right\|^2 - \sigma_n^2\log\det\left(\mathbb{P}_D^{\perp}\right) \\ &= \arg\min_{\mathbb{B}_D}\left\|\mathbb{P}_D^{\perp}\boldsymbol{y}_D\right\|^2,\end{aligned}$$

where $\mathbb{P}_D^{\perp} = \mathbb{I} - \mathbb{B}_D(\mathbb{B}_D^{\mathtt{H}}\mathbb{B}_D)^{-1}\mathbb{B}_D^{\mathtt{H}}$. Note that it can be shown that the determinant of $\mathbb{P}_D^{\perp}$ is always equal to 0. If however, the receiver has perfect CSI, then $E[\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}^{\mathtt{H}}]$ becomes an all-zero matrix. Hence, $\mathbb{C} = \sigma_n^2\mathbb{I}$, which reduces (4) to

$$\hat{\boldsymbol{b}}_{\text{ML, perfect CSI}} = \arg\min_{\mathbb{B}_D}\left\{\left\|\boldsymbol{y}_D - \mathbb{B}_D\hat{\boldsymbol{h}}\right\|^2\right\}. \tag{5}$$

B. near-ML decoding criterion at medium to high SNRs

At medium to high SNRs, the last term in (4) becomes negligible when being compared the first two terms because $\sigma_n^2 \approx 0$. We can then yield a near-ML decoding criterion as follows.

$$\begin{aligned}\hat{\boldsymbol{b}}_{\text{Near-ML}} &= \arg\min_{\mathbb{B}_D} \Big\{ \|\boldsymbol{y}_D - \mathbb{B}_D \hat{\boldsymbol{h}}\|^2 \\ &\quad - \left(\boldsymbol{y}_D - \mathbb{B}_D \hat{\boldsymbol{h}}\right)^{\mathtt{H}} \mathbb{Q}_B \left(\boldsymbol{y}_D - \mathbb{B}_D \hat{\boldsymbol{h}}\right) \Big\}. \quad (6)\end{aligned}$$

By [1], we know that the joint ML decoding of both the code and training sequence for given $\boldsymbol{y}$ is

$$\hat{\boldsymbol{b}}_{\text{JML}} = \arg\min_{\mathbb{B}_D} \left\{ \|\mathbb{P}_B^{\perp} \boldsymbol{y}\|^2 \right\},$$

where

$$\mathbb{P}_B^{\perp} = \mathbb{I} - \mathbb{P}_B = \mathbb{I} - \mathbb{B}(\mathbb{B}^{\mathtt{H}}\mathbb{B})^{-1}\mathbb{B}^{\mathtt{H}}.$$

We then found that the near-ML decoding criterion in (6) performs exactly the joint ML decoding in [1]. We summarize this result in the next lemma.

Lemma 1:

$$\hat{\boldsymbol{b}}_{\text{Near-ML}} = \hat{\boldsymbol{b}}_{\text{JML}}.$$

Proof: First, we note that the near-ML decoding criterion in (6) can be equivalently changed to:

$$\hat{\boldsymbol{b}}_{\text{Near-ML}} = \arg\min_{\mathbb{B}_D} \left\{ \left(\boldsymbol{y} - \mathbb{B}\hat{\boldsymbol{h}}\right)^{\mathtt{H}} \left(\mathbb{I} - \overline{\mathbb{Q}}_B\right) \left(\boldsymbol{y} - \mathbb{B}\hat{\boldsymbol{h}}\right) \right\}, \quad (7)$$

where

$$\overline{\mathbb{Q}}_B = \begin{bmatrix} \mathbf{0}_{T\times T} & \mathbf{0}_{T\times(L-T)} \\ \mathbf{0}_{(L-T)\times T} & \mathbb{Q}_B \end{bmatrix}.$$

Letting $\widetilde{\mathbb{Q}}_B = \mathbb{P}_B - \overline{\mathbb{Q}}_B$, we obtain that

$$\widetilde{\mathbb{Q}}_B = \begin{bmatrix} \mathbb{B}_P(\mathbb{B}^{\mathtt{H}}\mathbb{B})^{-1}\mathbb{B}_P^{\mathtt{H}} & \mathbb{B}_P(\mathbb{B}^{\mathtt{H}}\mathbb{B})^{-1}\mathbb{B}_D^{\mathtt{H}} \\ \mathbb{B}_D(\mathbb{B}^{\mathtt{H}}\mathbb{B})^{-1}\mathbb{B}_P^{\mathtt{H}} & \mathbf{0}_{(L-T)\times(L-T)} \end{bmatrix}.$$

Continuing the derivation in (7) yields

$$\begin{aligned}\hat{\boldsymbol{b}}_{\text{Near-ML}} &= \left(\boldsymbol{y} - \mathbb{B}\hat{\boldsymbol{h}}\right)^{\mathtt{H}} \left(\mathbb{P}_B^{\perp} + \widetilde{\mathbb{Q}}_B\right) \left(\boldsymbol{y} - \mathbb{B}\hat{\boldsymbol{h}}\right) \\ &= \boldsymbol{y}^{\mathtt{H}}\mathbb{P}_B^{\perp}\boldsymbol{y} + \boldsymbol{y}^{\mathtt{H}}\widetilde{\mathbb{Q}}_B\boldsymbol{y} - \hat{\boldsymbol{h}}^{\mathtt{H}}\mathbb{B}^{\mathtt{H}}\widetilde{\mathbb{Q}}_B\boldsymbol{y} - \boldsymbol{y}^{\mathtt{H}}\widetilde{\mathbb{Q}}_B\mathbb{B}\hat{\boldsymbol{h}} \\ &\quad + \hat{\boldsymbol{h}}^{\mathtt{H}}\mathbb{B}^{\mathtt{H}}\widetilde{\mathbb{Q}}_B\mathbb{B}\hat{\boldsymbol{h}} \\ &= \|\mathbb{P}_B^{\perp}\boldsymbol{y}\|^2,\end{aligned}$$

where the second equality follows because $\boldsymbol{y}^{\mathtt{H}}\widetilde{\mathbb{Q}}_B\boldsymbol{y} - \hat{\boldsymbol{h}}^{\mathtt{H}}\mathbb{B}^{\mathtt{H}}\widetilde{\mathbb{Q}}_B\boldsymbol{y} - \boldsymbol{y}^{\mathtt{H}}\widetilde{\mathbb{Q}}_B\mathbb{B}\hat{\boldsymbol{h}} + \hat{\boldsymbol{h}}^{\mathtt{H}}\mathbb{B}^{\mathtt{H}}\widetilde{\mathbb{Q}}_B\mathbb{B}\hat{\boldsymbol{h}} = 0$. ■

We end this section by remarking that if $\mathbb{B}_D^{\mathtt{H}}\mathbb{B}_D$ is equal to a constant, then it can be derived from (4) that $\hat{\boldsymbol{b}}_{\text{ML}} = \hat{\boldsymbol{b}}_{\text{Near-ML}} = \hat{\boldsymbol{b}}_{\text{JML}}$. As a result, the ML decoding criterion in (4) is simplified to the near-ML (equivalently, JML) decoding criterion.

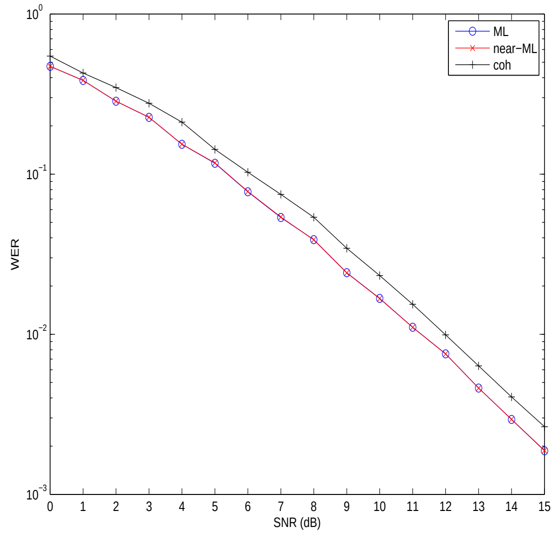

Fig. 1. Performances of the ML decoding criterion in (4) with perfect knowledge of σ_n^2 (ML), the near-ML decoding criterion in (6) (near-ML), and the coherent receiver using $\hat{\boldsymbol{h}}$ as the perfect-CSI and adopting the decoding criterion in (5) (coh).

IV. Iterative Joint Code and Training Sequence Design Algorithm

In this section, we provide an iterative algorithm to design the code and training sequence jointly.

The cost function f that is to minimize during the search process is the union bound constituted by pairwise error probabilities (PEPs)

$$f(\mathcal{C}, \mathbb{B}_P) = \sum_{i,j} \Pr\{\mathbb{B}_D(j) \text{ claimed}|\mathbb{B}_D(i) \text{ transmitted}, \mathbb{B}_P\} \quad (8)$$

where $\mathcal{C} = \{\mathbb{B}_D(0), \cdots, \mathbb{B}_D(2^K - 1)\}$, and K is the length of information bits. Since the near-ML decoding criterion is equivalent to the JML decoding criterion, (8) can be well approximated by the PEP union bound derived for the JML decoding in [1] at medium to high SNRs.

Thus, after an initial training sequence $\mathbb{B}_P^{(0)}$ is generated, simulated annealing algorithm [1] is employed to search for a code $\mathcal{C}^{(0)}$ that has an acceptably small PEP union bound. Based on the $\mathcal{C}^{(0)}$, the next $\mathbb{B}_P^{(1)}$ is again obtained via simulated annealing. The process is repeated until the maximum number of iterations is achieved. Detail is summarized below.

Step 1. Set the maximum iteration number I_M and initialize $i = 0$. Generate $\mathbb{B}_P^{(i)}$ randomly.

Step 2. Perform simulated annealing algorithm *to search for a good code $\mathcal{C}^{(i)}$ based on the cost function $f(\mathcal{C}^{(i)}, \mathbb{B}_P^{(i)})$.*

Step 3. Set $i = i + 1$. Perform simulated annealing algorithm *to search for a good training sequence $\mathbb{B}_P^{(i)}$ based on the cost function $f(\mathcal{C}^{(i-1)}, \mathbb{B}_P^{(i)})$.*

Step 4. If $i > I_M$, output $\mathbb{B}_P^{(i)}$ and $\mathcal{C}^{(i)}$ and stop the algorithm; otherwise, go to Step 2.

V. Simulations

Same as in [1], the channel coefficients $\boldsymbol{h}$ used in our simulations is zero-mean complex-Gaussian distributed with $E[\boldsymbol{h}\boldsymbol{h}^{\mathrm{H}}] = (1/P)\mathbb{I}$ and $P = 2$. The SNR of the system is given by

$$\mathrm{SNR} = \frac{\mathrm{tr}\left(E[\boldsymbol{h}\boldsymbol{h}^{\mathrm{H}}]\right)}{\sigma_n^2} = \frac{E[\boldsymbol{h}^{\mathrm{H}}\boldsymbol{h}]}{\sigma_n^2} = \frac{1}{\sigma_n^2}.$$

We then examined the system consisting of the 7-bit training sequence $[b_1, \ldots, b_7] = [0,0,0,0,1,0,1]$ and the $(15,11)$ Hamming code. The performances of receivers respectively using the ML decoding criteron in (4) with perfect knowledge of noise power σ_n^2, and the near-ML decoding criterion in (6), as well as the coherent receiver using $\hat{\boldsymbol{h}}$ as the perfect-CSI and adopting the decoding criterion in (5), are summarized in Fig. 1. The results show that the ML decoding outperforms the coherent receiver, and the performance gap is around 0.7 dB. Furthermore, the near-ML decoding, which requires no knowledge of σ_n^2, performs as good as the ML decoding. We conclude from Fig. 1 that the near-ML criterion, as we have claimed by its naming, can yield almost the same performance as the ML criterion at medium to high SNRs. As a result, our proposed joint design of code and training sequence based on the near-ML criterion is then justified.

We next compare the performance of the joint design of code and training sequence obtained from our iterative search algorithm with that of the $(7,4)$ Hamming code and training sequence $[b_1, \ldots, b_4] = [0,1,0,0]$ as shown in Fig. 2. The code our algorithm found is $\{10,\ 13,\ 16,\ 23,\ 33,\ 38,\ 59,\ 60,\ 67,\ 68,\ 88,\ 95,\ 105,\ 110,\ 114,\ 117\}$, where the numbers in the parentheses are calculated based on the formula $\sum_{j=1}^{N-T}\left(b_{T+j}\cdot 2^{N-T-j}\right)$. Our results show that our code can further improve the ML performance of the $(7,4)$ Hamming code and training sequence $[b_1, \ldots, b_4] = [0,1,0,0]$ with a gain about 0.3 dB.

By examining codes of different code rates further (for which the results are not shown in the paper due to the page limits), we found that more performance gain can be obtained by our joint design when a smaller code rate is concerned. As an example, when the extreme case with $K = 1$ (i.e., there are $2^K = 2$ codewords) is considered, Fig. 3 shows that our proposed joint design can considerably outperform the channel code that consists of all-$(+1)$ and all-(-1) codewords, which we term "antipolar" in the figure. The code our algorithm found is $\{221, 1826\}$ in decimal representation. Note that the antipolar channel code we compare our design with should be optimal in performance when perfect channel estimation is assumed.

VI. Conclusion

After confirming the well approximation of the near-ML criterion to the ML criterion at medium to high SNRs, we proposed in this work an iterative algorithm to jointly design the code and training sequence for a receiver to have partial CSI information based on the near-ML criterion. Simulations

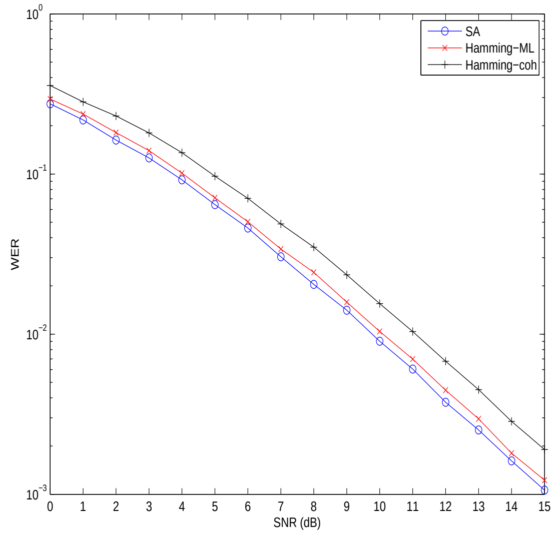

Fig. 2. ML performances of the $(7,4)$ code and 4-bit training sequence $[b_1, \ldots, b_4] = [1,0,1,1]$ found by our simulated-annealing-based iterative search algorithm (SA), and the (7,4)-Hamming code and training sequence $[b_1, \ldots, b_4] = [0,1,0,0]$ (Hamming-ML). Also shown is the performance of the coherent receiver for the (7,4)-Hamming code and training sequence $[b_1, \ldots, b_4] = [0,1,0,0]$ using $\hat{\boldsymbol{h}}$ as the perfect-CSI and adopting the decoding criterion in (5) (Hamming-coh).

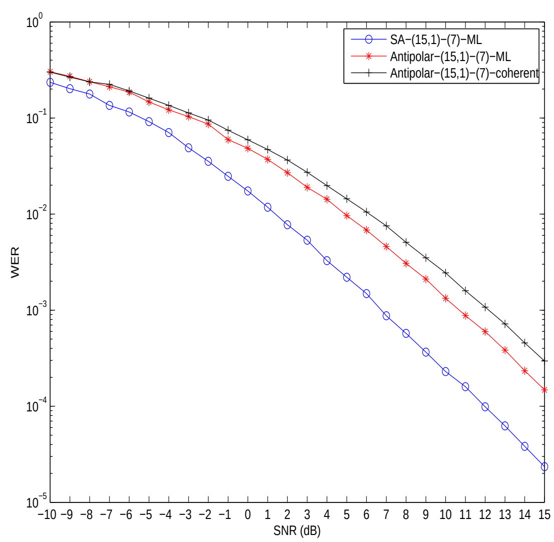

Fig. 3. ML performances of the $(15,1)$ code and 7-bit training sequence $[b_1, \ldots, b_7] = [1,0,1,0,0,1,1]$ found by our simulated-annealing-based iterative search algorithm (SA-(15,1)-(7)-ML), and the (15,1)-binary code and training sequence $[b_1, \ldots, b_7] = [0,0,0,0,1,0,1]$ (Antipolar-(15,1)-(7)-ML). Also shown is the performance of the coherent receiver for the (15,1) code and training sequence $[b_1, \ldots, b_7] = [0,0,0,0,1,0,1]$ using $\hat{\boldsymbol{h}}$ as the perfect-CSI and adopting the decoding criterion in (5) (Antipolar-(15,1)-(7)-coherent)

show that the code and training sequence that are found by our iterative algorithm can outperform the traditional system with a channel code and a training sequence satisfying the self-orthogonal condition $\mathbb{B}_P^{\mathtt{H}}\mathbb{B}_P = T\mathbb{I}$, particularly when codes of low rate are considered.

REFERENCES

[1] M. Skoglund, J. Giese and S. Parkvall, "Code design for combined channel estimation and error protection," *IEEE Trans. Inform. Theory*, vol. 48, no. 5, pp. 1162-1171, May 2002.

[2] O. Coskun and K. M. Chugg, "Combined coding and training for unknown ISI channels," *IEEE Trans. Commun.*, vol. 53, no. 8, pp. 1310-1322, August 2005.

[3] C.-L. Wu, P.-N. Chen, Yunghsiang S. Han and M.-H. Kuo, "Maximum-likelihood priority-first search decodable codes for combined channel estimation and error correction," *IEEE Trans. Inform. Theory*, vol. 55, no. 9, Sept 2009.

[4] M. Beko, J. Xavier and V. A. N. Barroso, "Noncoherent communication in multiple-antenna system: receiver design and codebook construction," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5703-5715, Dec. 2007.

[5] C. Fragouli, N. Al-Dhahir and W. Turin, "Finite-Alphabet Constant-Amplitude Training Sequences for Multiple-Antennas," *Proc. IEEE International Conference on Communications (ICC 2002)*, New York, USA, April 28-May 2, 2002.

[6] A. Viterbi, "Optimum detection and signal selection for partially coherent binary communication," *IEEE Trans. Inform. Theory*, vol. 11, no. 2, pp. 239-246, April 1965.

[7] M.K. Simon and D. Divsalar,"Multiple symbol partially coherent detection of MPSK," *IEEE Trans. Commun.*, vol. 42, no. 2/3/4, pp. 430-439, Feb/Mar/Apr 1994.

[8] K. Gomadam and S. A. Jafar, "Modulation and Detection for simple receivers in rapidly time varying channels," *IEEE Trans. Commun.*, vol. 55, no. 3, pp. 529-539, March 2007.

[9] B. Zhang and K. Kiasaleh, "Partially-Coherent Receiver Architectures for QAM Communications in the Presence of Non-constant Phase Estimation Error," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 568-573, February 2009.

[10] M.J. Borran, A. Sabharwal and B. Aazhang, "Design Criterion and Construction Methods for Partially Coherent Multiple Antenna Constellations," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4122-4133, February 2009.

[11] J. Giese and M. Skoglund, "Space-Time Constellation Design for Partial CSI at the Receiver," *IEEE Trans. Inform. Theory*, vol. 53, no. 8, pp. 2715-2731, August 2007.