

Superposition Coded Modulation and Iterative Linear MMSE Detection

Li Ping, *Senior Member, IEEE*, Jun Tong, Xiaojun Yuan, *Member, IEEE*, and Qinghua Guo, *Member, IEEE*

Abstract—We study superposition coded modulation (SCM) with iterative linear minimum-mean-square-error (LMMSE) detection. We show that SCM offers an attractive solution for highly complicated transmission environments with severe interference. We analyze the impact of signaling schemes on the performance of iterative LMMSE detection. We prove that among all possible signaling methods, SCM maximizes the output signal-to-noise/interference ratio (SNIR) in the LMMSE estimates during iterative detection. Numerical examples are used to demonstrate that SCM outperforms other signaling methods when iterative LMMSE detection is applied to multi-user/multi-antenna/multi-path channels.

Index Terms—Iterative detection, minimum mean square error (MMSE), superposition coded modulation (SCM).

I. INTRODUCTION

CODED modulation can deliver high-rate, reliable transmissions over noisy channels. The early work on trellis coded modulation (TCM) [1] and multilevel codes [2] demonstrates that significant performance gains are achievable by properly combining binary coding and multi-ary signaling. Bit interleaved coded modulation (BICM) [3] provides an even simpler solution. After the advent of turbo and low-density parity-check (LDPC) codes [4], [5], iterative decoding has been widely studied to improve the performance of coded modulation schemes [6]-[9]. In particular, BICM with iterative decoding (BICM-ID) [10], [11] has emerged as an attractive option due to its simplicity and good performance in both fading and additive white Gaussian noise (AWGN) channels.

In a simple AWGN channel, the detection complexity is $O(2^Q)$ for a conventional 2^Q -ary coded modulation scheme carrying rQ information bits per symbol, where $r \leq 1$ represents redundancy introduced by coding. Such exponential complexity constitutes a major obstacle when rate is high.

The complexity problem becomes more severe for channels with interference. For example, consider a multi-user/multi-path channel with K users and L paths. Let the signal at each transmit antenna of a user be drawn from a constellation of size 2^Q . At the receiver, the signals from different users/paths are superimposed to form a joint constellation of size 2^{QKL} .

Manuscript received 1 October 2008; revised 16 January 2009. This work was fully supported by a grant from the Research Grant Council of the Hong Kong Special Administrative Region, China, under project CityU 116706.

Li Ping and Jun Tong are with the Department of Electronic Engineering, City University of Hong Kong, Tat Chee Avenue, Kowloon Tong, Kowloon, Hong Kong (e-mail: eeliping@cityu.edu.hk, eejtong@gmail.com).

Xiaojun Yuan is with the Department of Electronic Engineering, University of Hawaii at Manoa, Honolulu, HI 96822, USA (e-mail: xyuan24@hawaii.edu).

Qinghua Guo is with the School of Electrical, Electronic and Computer Engineering, The University of Western Australia, WA, Australia (e-mail: qhguo@ee.uwa.edu.au).

Digital Object Identifier 10.1109/JSAC.2009.090817.

A direct approach needs to compute the metric between the received symbol and every point in the constellation. The related complexity $O(2^{QKL})$ can be prohibitively high even for modest Q , K and L values.

This paper is concerned with superposition coded modulation (SCM) [12]-[20] that can be regarded as a special case of BICM. Our emphasis is on iterative linear minimum mean square error (linear MMSE or LMMSE) detection in complicated channel conditions [21]-[27]. In this case, SCM offers an attractive solution as its detection complexity can be made as low as $O(\log J)$ per bit using the fast Fourier transform (FFT)-based technique proposed in [24] (see also the discussions in Section II-C and III-C below), where J is the length of a frame. This complexity is not a function of Q , K and L in a multi-user/multi-path channel and so it is significantly lower than the direct approach mentioned above.

The main contribution of this paper is a proof that among all possible signaling methods, SCM maximizes the output signal-to-noise/interference ratio (SNIR) (or equivalently, minimizes the output mean square error (MSE)) in the LMMSE estimates during iterative detection. Thus, SCM can potentially outperform other alternatives when iterative LMMSE detection is employed. This provides a useful guideline for signaling design in complicated transmission environments. Numerical results are provided to verify the theoretical analysis.

The finding in this paper also provides a theoretical explanation to the observation that signaling methods (including both constellation patterns and mapping rules) have a strong impact on the performance of iterative LMMSE detection [26], [27]. It is observed in [26] that the optimized signaling methods for maximum *a posteriori* (MAP) detectors may not be good choices for iterative LMMSE detectors. However, the work in [26], [27] is mostly based on simulations. To the best of our knowledge, this paper presents the first rigorous analytical study on this issue.

II. SUPERPOSITION CODED MODULATION

This section provides a brief overview of SCM and the related transmission principles in a simple AWGN channel. The treatment in more complicated channels will be discussed in the next section.

A. SCM Signaling

A signaling scheme is characterized by a constellation \mathcal{S} and a mapping rule \mathcal{R} . In a conventional scheme, these two concepts can be defined separately; the former refers to the placement of signal points and the latter refers to indexing the signal points with binary labels. For SCM, however, the two concepts are closely related as shown below. For

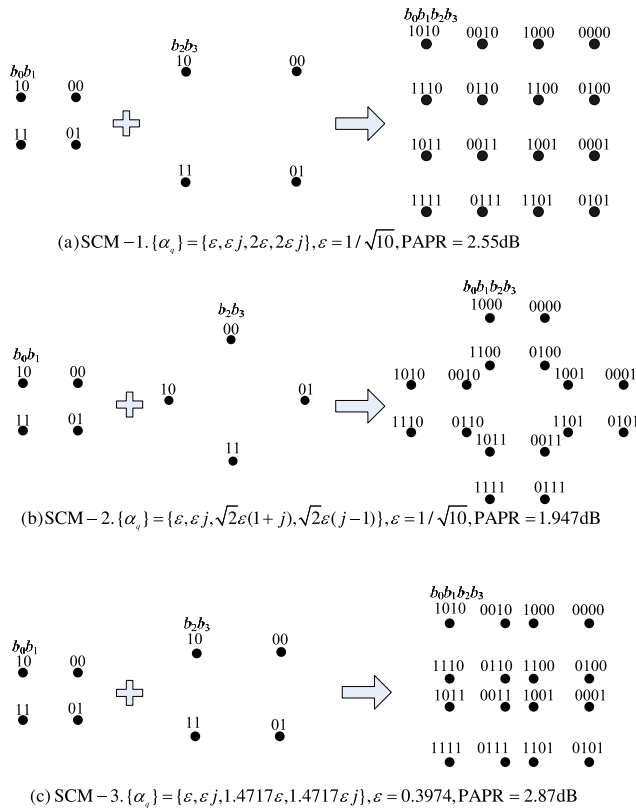


Fig. 1. Three 16-ary SCM signaling schemes formed by different $\{\alpha_q\}$ sets, where ε is a normalization factor to meet the constraint in (3) and $j = \sqrt{-1}$.

this reason, we will use the terms ‘‘SCM constellation’’ and ‘‘SCM mapping’’. (Note that SCM mapping is also referred to as sigma mapping in [14], [17] and interleave-division multiplexing (IDM) in [15].)

Definition 1: Given Q arbitrary complex coefficients $\{\alpha_q\}$ and a set $\mathcal{B} = \{B \equiv \{b_q\}\}$ of binary Q -tuples, the SCM mapping $\mathcal{R} : \mathcal{B} \rightarrow \mathcal{S}$ is defined as

$$s_B = \sum_{q=0}^{Q-1} \alpha_q (-1)^{b_q}. \quad (1)$$

An SCM constellation $\mathcal{S} = \{s_B\}$ is formed by running (1) over 2^Q possible values of B .

For example, Gray-mapped quadrature phase shift keying (QPSK) is a $Q = 2$ SCM scheme with $s_B = (-1)^{b_0} + j(-1)^{b_1}$, i.e., $\alpha_0 = 1$ and $\alpha_1 = j \equiv \sqrt{-1}$. For an even Q , we can construct SCM by superimposing several Gray-mapped QPSK constellations. Fig. 1 illustrates three examples of such SCM schemes with $Q = 4$ and different $\{\alpha_q\}$ sets. The mappings are shown by the binary labeling $\{b_q\} = b_0b_1b_2b_3$. For convenience, they are referred to as SCM-1, SCM-2 and SCM-3, respectively. They have different properties as follows and their performance is demonstrated in Section V.

- SCM-1 is a standard 16-QAM constellation, but the other two are not.
- SCM-2 has the lowest peak-to-average-power ratio (PAPR) among the three. It performs slightly poorer than SCM-1 since the minimum distance between signal points is smaller in SCM-2.
- SCM-3 is optimized for a turbo-coded scheme. See Fig. 5 below.

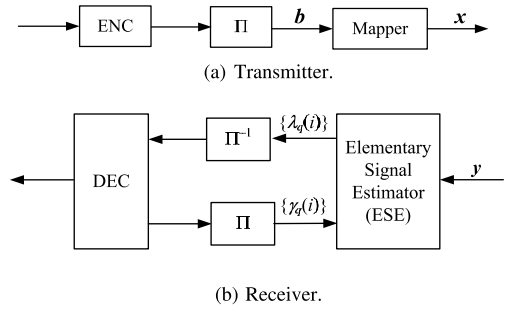


Fig. 2. Transmitter and receiver structures of an SCM system. ENC and DEC denote the encoder and decoder, respectively. Π denotes the interleaver and Π^{-1} the corresponding de-interleaver.

SCM has the following notable features.

- 1) SCM can be detected at a low cost. See Section II-C and III-F below.
- 2) SCM is convenient for adaptive modulation [15]. The transmitter and receiver structures remain roughly the same for different Q .
- 3) SCM provides a remedy to the PAPR problem in orthogonal frequency-division multiplexing (OFDM). The related discussion can be found in [19].
- 4) It is easy to optimize SCM, since it involves only Q parameters $\{\alpha_q\}$ for a 2^Q -ary constellation. In contrast, the optimization for a general BICM over a 2^Q -ary constellation has complexity $O(2^Q!)$ with exhaustive search.
- 5) SCM offers improved performance when LMMSE detection is used. This is the focus of the present paper.
- 6) The linear superposition of several SCM signals naturally results in an SCM signal. These signals can be encoded distributively at different users/antennas/relay nodes. The receiver structure remains roughly the same for different variations. This property is very useful in multiple access/relay/ad hoc networks, where a challenging issue is how to separate signals from different sources after they are superimposed. SCM provides a simple solution to this issue, following its superposition nature and property 1 above.
- 7) SCM can be used to construct space-time codes that can achieve high diversity and multiplexing gains in MIMO systems [28].

B. SCM over an AWGN Channel

Fig. 2 shows the transmitter and receiver structures for SCM. They are basically the same as for BICM-ID, except that the constellation involved can be ‘‘unconventional’’ as in Figs. 1(b) and 1(c). At the transmitter, the information data is first encoded by a binary forward error-control (FEC) code. The interleaved coded bits are grouped into binary Q -tuple:

$$\mathbf{b}(i) \equiv \{b_0(i), b_1(i), \dots, b_{Q-1}(i)\}. \quad (2)$$

The mapper then maps each $\mathbf{b}(i)$ to a symbol $x(i) \in \mathcal{S}$ based on (1). We naturally assume that each $b_q(i)$ is equally taken over $\{0, 1\}$. We assume that \mathcal{S} is unbiased and with normalized power, i.e.,

$$\sum_{s \in \mathcal{S}} s = 0, \quad \text{and} \quad \frac{1}{2^Q} \sum_{s \in \mathcal{S}} |s|^2 = 1. \quad (3)$$

The signal sequence $\mathbf{x} = \{x(i)\}$ is transmitted over the channel, yielding the received signal sequence $\mathbf{y} = \{y(i)\}$. The iterative receiver as shown in Fig. 2(b) follows the general turbo principle [4]. The decoding function of the decoder (DEC) is standard; therefore we focus only on the elementary signal estimator (ESE).

For an AWGN channel, a received symbol is given by

$$y(i) = x(i) + w(i), \quad (4)$$

where $w(i)$ is the channel noise sample with mean zero and variance $\sigma^2 = N_0/2$ per dimension. With *a posteriori* probability (APP) de-mapping, the ESE output is the log-likelihood ratio (LLR) for $b_q(i)$,

$$\begin{aligned} v_q(i) &\equiv \ln \left(\frac{\sum_{s \in \mathcal{S}_q^0} \Pr(y(i)|s) \Pr(s)}{\sum_{s \in \mathcal{S}_q^1} \Pr(y(i)|s) \Pr(s)} \right) \\ &= \ln \left(\frac{\sum_{s \in \mathcal{S}_q^0} \exp\left(-\frac{|y(i)-s|^2}{N_0}\right) \Pr(s)}{\sum_{s \in \mathcal{S}_q^1} \exp\left(-\frac{|y(i)-s|^2}{N_0}\right) \Pr(s)} \right), \quad (5) \end{aligned}$$

where \mathcal{S}_q^0 and \mathcal{S}_q^1 are the subsets of the constellation points in \mathcal{S} whose q th bit carries 0 and 1, respectively. In (5), $\Pr(s)$ is the *a priori* probability for $x(i) = s$. It can be computed from the DEC feedback $\{\gamma_q(i)\}$ (the *extrinsic* LLR about $\{b_q(i)\}$ from the DEC), as detailed later in Section IV-A. We also compute the so-called *extrinsic* LLR as follows.

$$\lambda_q(i) \equiv v_q(i) - \gamma_q(i). \quad (6)$$

In the iterative detection process, the *extrinsic* LLRs are delivered to the DEC, following the turbo principle [4]. The ESE operation defined in (5) has complexity $O(2^Q/Q)$ per bit¹.

The above is for an AWGN channel. Detection for more general environments such as multi-user, multiple-input multiple-output (MIMO) and multi-path channels will be discussed in Section III.

C. Layer-by-Layer De-mapping

For SCM, a layer-by-layer de-mapping method can be used at the receiver based on a Gaussian approximation as follows. Focusing on a particular bit $b_q(i)$, we rewrite (4) as

$$y(i) = \alpha_q(-1)^{b_q(i)} + \zeta_q(i), \quad (7)$$

where

$$\zeta_q(i) = \sum_{q'=0, q' \neq q}^{Q-1} \alpha_{q'}(-1)^{b_{q'}(i)} + w(i) \quad (8)$$

denotes the interference/noise component with respect to $b_q(i)$. We approximate $\zeta_q(i)$ using a Gaussian random variable. Then (7) can be seen as a simple binary-input scheme. In the iterative decoding process, the *extrinsic* information from the DEC can be used to estimate the mean and variance of

¹The complexity is $O(2^Q)$ by applying (5) directly to every bit. However, some operations in (5) can be shared for different bits, then the complexity can be reduced to $O(2^Q/Q)$ per bit. The detailed algorithm is equivalent to the fast Hadamard transform discussed in [37].

$\zeta_q(i)$. The ESE output can be computed based on the Gaussian assumption as in (9) at the top of the next page, where $\text{Re}(\cdot)$ denotes the real part of a complex number and $(\cdot)^*$ denotes the conjugate operation. The details for finding $E(\text{Re}(\alpha_q^* \zeta_q(i)))$ and $\text{Var}(\text{Re}(\alpha_q^* \zeta_q(i)))$ can be found in [19]. Interestingly, the complexity in (9) is $O(1)$ per coded bit, independent of the constellation size.

III. GENERAL LINEAR CHANNELS AND ITERATIVE LMMSE DETECTION

A. Generic Linear Model and Gaussian Approximation

Consider a generic linear system:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\eta}, \quad (10)$$

where \mathbf{y} and $\mathbf{x} = \{x(i)\}$ are the received and transmitted signal vectors, respectively, and $\boldsymbol{\eta}$ is a vector of AWGN with covariance matrix $N_0\mathbf{I}$. In the above, \mathbf{H} represents the multiplicative effect of the channel, and \mathbf{x} is generally a segment in a coded sequence. (See Section V-B.) We always assume that \mathbf{H} is known perfectly at the receiver. The applications of the model in (10) include multi-user, MIMO and multi-path channels or their combinations, as discussed in [24].

In general, the optimal MAP [33] solution to the ESE can be very complicated for the system in (10). The following sub-optimal LMMSE approach [20], [21] is a low-cost alternative. It consists of three functional modules as shown in Fig. 3. They are briefly explained below.

Gaussian Mapper:

We generate the mean and covariance matrix of \mathbf{x} , denoted by $\boldsymbol{\mu}_x$ and \mathbf{V}_x respectively, using the DEC feedback. (We will discuss this issue in detail in Section IV-A.) We assume that \mathbf{x} is Gaussian and can be fully characterized by $\boldsymbol{\mu}_x$ and \mathbf{V}_x . (Note: This is only an approximation since \mathbf{x} is actually discrete with coded modulation.) We also assume that the entries of \mathbf{x} are independent of each other which can be approximately ensured by random interleaving) and so \mathbf{V}_x is a diagonal matrix, i.e.,

$$\mathbf{V}_x = \text{diag}(\text{Var}(x(0)), \text{Var}(x(1)), \dots, \text{Var}(x(J-1))), \quad (11)$$

where J is the length of \mathbf{x} .

LMMSE Estimator:

Based on the above Gaussian approximation, the MMSE estimate of \mathbf{x} is [30]

$$\hat{\mathbf{x}} \equiv E(\mathbf{x}|\mathbf{y}) = \boldsymbol{\mu}_x + \mathbf{V}_x \mathbf{H}^H \mathbf{R}^{-1} (\mathbf{y} - E(\mathbf{y})), \quad (12)$$

where $E(\mathbf{y}) \equiv \mathbf{H}\boldsymbol{\mu}_x$, $(\cdot)^H$ denotes the conjugate transpose operation, and

$$\mathbf{R} \equiv E((\mathbf{y} - E(\mathbf{y}))(\mathbf{y} - E(\mathbf{y}))^H) = \mathbf{H}\mathbf{V}_x\mathbf{H}^H + N_0\mathbf{I} \quad (13)$$

is the covariance matrix of \mathbf{y} . The complexity in (12) is dominated by the matrix inversion \mathbf{R}^{-1} .

Soft De-mapper:

Denote by $\mathbf{h}(i)$ the i th column of \mathbf{H} . We write the i th entry $\hat{x}(i)$ of $\hat{\mathbf{x}}$ in (12) in a signal plus interference-noise form as [21]:

$$\hat{x}(i) = \phi(i)x(i) + \xi(i), \quad (14)$$

where

$$\phi(i) \equiv \text{Var}(x(i))\mathbf{h}(i)^H \mathbf{R}^{-1} \mathbf{h}(i) \quad (15)$$

$$\begin{aligned}
v_q(i) &= \ln \left(\frac{\exp \left(-\frac{|\operatorname{Re}(\alpha_q^* y(i)) - |\alpha_q|^2 - \operatorname{E}(\operatorname{Re}(\alpha_q^* \zeta_q(i)))|^2}{2\operatorname{Var}(\operatorname{Re}(\alpha_q^* \zeta_q(i)))} \right) \Pr(b_q(i)=0)}{\exp \left(-\frac{|\operatorname{Re}(\alpha_q^* y(i)) + |\alpha_q|^2 - \operatorname{E}(\operatorname{Re}(\alpha_q^* \zeta_q(i)))|^2}{2\operatorname{Var}(\operatorname{Re}(\alpha_q^* \zeta_q(i)))} \right) \Pr(b_q(i)=1)} \right) \\
&= 2 \frac{|\alpha_q|^2 (\operatorname{Re}(\alpha_q^* y(i)) - \operatorname{E}(\operatorname{Re}(\alpha_q^* \zeta_q(i))))}{\operatorname{Var}(\operatorname{Re}(\alpha_q^* \zeta_q(i)))} + \gamma_q(i)
\end{aligned} \tag{9}$$

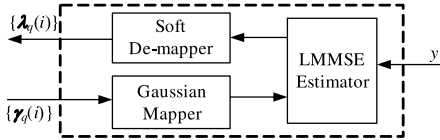


Fig. 3. LMMSE approach to the ESE.

$$\xi(i) \equiv \operatorname{E}(x(i)) + \operatorname{Var}(x(i)) \mathbf{h}(i)^H \mathbf{R}^{-1} (\mathbf{y} - \operatorname{E}(\mathbf{y}) - \mathbf{h}(i)x(i)). \tag{16}$$

We approximate $\xi(i)$ as an additive Gaussian noise independent of $x(i)$. The APP de-mapping rule defined in (5) can now be applied to (14) with complexity $O(2^Q/Q)$ per bit:

$$v_q(i) \equiv \ln \left(\frac{\sum_{s \in \mathcal{S}_q^0} \exp \left(-\frac{|\hat{x}(i) - \phi(i)s - \operatorname{E}(\xi(i))|^2}{\operatorname{Var}(\xi(i))} \right) \Pr(s)}{\sum_{s \in \mathcal{S}_q^1} \exp \left(-\frac{|\hat{x}(i) - \phi(i)s - \operatorname{E}(\xi(i))|^2}{\operatorname{Var}(\xi(i))} \right) \Pr(s)} \right). \tag{17}$$

Then, the *extrinsic* LLR $\lambda_q(i)$ can be calculated using (6). Alternatively, the layer-by-layer de-mapping technique similar to that in (9) can be used with complexity $O(1)$ per bit. We omit the details for space limitations.

Intuitively, in the above, the Gaussian mapper generates the *a priori* distribution of \mathbf{x} (characterized by $\boldsymbol{\mu}_x$ and \mathbf{V}_x) under the Gaussian assumption. The LMMSE estimator generates $\hat{\mathbf{x}}$ that is the *a posteriori* mean of \mathbf{x} after observing \mathbf{y} . Statistically, $\hat{\mathbf{x}}$ is the closest to \mathbf{x} under the Gaussian assumption. The soft de-mapper estimates each bit according to the discrete distribution of each individual $x(i)$ but treats interference as Gaussian noise. This is sub-optimal, but the problem is greatly simplified.

B. SNIR in the LMMSE Estimates

The de-mapping performance is determined by the SNIR in the signal model (14) with respect to the desired signal $x(i)$, which can be evaluated as follows. From (16), $\operatorname{E}(\xi(i)) = (1 - \phi(i))\operatorname{E}(x(i))$, based on which it can be shown that

$$\operatorname{Var}(\xi(i)) = \phi(i)(1 - \phi(i))\operatorname{Var}(x(i)). \tag{18}$$

Assume that $\operatorname{E}(|x(i)|^2)$ is normalized to 1. The SNIR in (14) is given by

$$\begin{aligned}
SNIR &= \frac{|\phi(i)|^2 \operatorname{E}(|x(i)|^2)}{\operatorname{Var}(\xi(i))} = \frac{\phi(i)}{(1 - \phi(i))\operatorname{Var}(x(i))} \\
&= \frac{\mathbf{h}(i)^H \mathbf{R}^{-1} \mathbf{h}(i)}{1 - \operatorname{Var}(x(i)) \mathbf{h}(i)^H \mathbf{R}^{-1} \mathbf{h}(i)}.
\end{aligned} \tag{19}$$

Let $\mathbf{R}(i) = \sum_{i' \neq i} \operatorname{Var}(x(i')) \mathbf{h}(i') \mathbf{h}(i')^H + N_0 \mathbf{I} = \mathbf{R} - \operatorname{Var}(x(i)) \mathbf{h}(i) \mathbf{h}(i)^H$. From the matrix inversion lemma,

$$\mathbf{R}^{-1} = \mathbf{R}(i)^{-1} - \frac{\operatorname{Var}(x(i)) \mathbf{R}(i)^{-1} \mathbf{h}(i) \mathbf{h}(i)^H \mathbf{R}(i)^{-1}}{1 + \operatorname{Var}(x(i)) \mathbf{h}(i)^H \mathbf{R}(i)^{-1} \mathbf{h}(i)}.$$

Then from (19), it can be shown that

$$SNIR = \mathbf{h}(i)^H \mathbf{R}(i)^{-1} \mathbf{h}(i). \tag{20}$$

The SNIR in (20) is a monotonically decreasing function of $\operatorname{Var}(x(i'))$ for every $i' \neq i$, since

$$\begin{aligned}
\frac{dSNIR}{d\operatorname{Var}(x(i'))} &= \mathbf{h}(i)^H \frac{d\mathbf{R}(i)^{-1}}{d\operatorname{Var}(x(i'))} \mathbf{h}(i) \\
&= -\mathbf{h}(i)^H \mathbf{R}(i)^{-1} \mathbf{h}(i') \mathbf{h}(i')^H \mathbf{R}(i)^{-1} \mathbf{h}(i) \\
&\leq 0.
\end{aligned} \tag{21}$$

This means that minimizing $\{\operatorname{Var}(x(i))\}$ can maximize the SNIR in the LMMSE estimates and thus improve performance. This observation is one of the main motivations of the work presented in this paper. We will return to this in Section IV.

C. Complexity

Here is a rough estimate of the complexity for the LMMSE-based ESE outlined above. Let us consider a single-path MIMO channel with N transmit antennas and M receiver antennas. Then \mathbf{H} is an $M \times N$ matrix and $\mathbf{R} = \mathbf{H} \mathbf{V}_x \mathbf{H}^H + N_0 \mathbf{I}$ is an $M \times M$ matrix. The complexity involved in the matrix inversion \mathbf{R}^{-1} is $O(M^3)$. The overall complexity per symbol can be roughly estimated as $O(2^Q + M^3/N)$ (with $O(M^3/N)$ for evaluating (12) and $O(2^Q)$ for the Gaussian mapping and APP de-mapping). This is significantly lower than the complexity $O(2^{NQ}/N)$ for the optimal MAP detector. (In the above, we have assumed a BLAST-type scheme as will be considered in Section V-B.)

We can also use (10) to represent a multi-path MIMO channel. In this case, \mathbf{H} is an $MJ \times NJ$ matrix with a Toeplitz type structure. The detection complexity becomes $O(2^Q + (MJ)^3/(NJ))$ per symbol that can be very high. However, using a cyclic prefixing technique, \mathbf{H} can be converted into a circulant matrix. Then \mathbf{R}^{-1} can be computed using the FFT technique, with complexity reduced to $O(2^Q + (1 + M/N) \log J + M^2)$ per symbol. Interested readers are referred to [24] (where the treatments for general multi-user/multi-antenna/multi-path systems are discussed) for details.

D. Iterative Detection Principles

Recall that in Section III-A, we need to generate the *a priori* information $\boldsymbol{\mu}_x$ and \mathbf{V}_x to perform the LMMSE operation

in (12). In the iterative detection process, the DEC outputs $\{\gamma_q(i)\}$ are fed back to the Gaussian mapper to update the *a priori* information $\boldsymbol{\mu}_x$ and \mathbf{V}_x , as detailed in Section IV-A below. The ESE and DEC operations are executed again to refine the estimates. This process continues iteratively for a preset number of iterations. A hard decision is then made in the final iteration to produce the data estimates. Related discussions on the iterative detection process can be found in [21]-[27].

E. Variance of the Feedback Information

During the above iterative process, the LMMSE estimation in (12) refines the estimate of \mathbf{x} using observation \mathbf{y} together with the *a priori* information $\boldsymbol{\mu}_x$ and \mathbf{V}_x . According to Section III-B, a more accurate value for $\boldsymbol{\mu}_x$ will lead to a higher SNIR and so to a better estimate. The accuracy of $\boldsymbol{\mu}_x$ can be measured by $\text{Var}(x(i))$. Interestingly, $\text{Var}(x(i))$ is a function of the signaling scheme, as we will see in Section IV. This implies that the performance of the above iterative process depends on the signaling method used.

F. Layer-by-Layer De-mapping Again

For SCM characterized by (1), a transmit symbol s can be written as $s = \boldsymbol{\alpha}^T \mathbf{x}$, where $\boldsymbol{\alpha} = [\alpha_0, \alpha_1, \dots, \alpha_{Q-1}]^T$ is a weighting vector and $\mathbf{x} \equiv [x_0, x_1, \dots, x_{Q-1}]^T = [(-1)^{b_0}, (-1)^{b_1}, \dots, (-1)^{b_{Q-1}}]^T$ is a BPSK-modulated signal vector. The output symbol of an AWGN channel can be written as

$$y = \boldsymbol{\alpha}^T \mathbf{x} + \eta, \quad (22)$$

which can be seen as a special case of the generic model in (10). We now show that in this case the LMMSE detection in Section III-A reduces to the layer-by-layer de-mapping in Section II-C. First, for y above, (12) can be rewritten as

$$\hat{\mathbf{x}} \equiv \boldsymbol{\mu}_x + \frac{\mathbf{V}_x \boldsymbol{\alpha}^*}{\sum_{q=0}^{Q-1} |\alpha_q|^2 \text{Var}(x_q) + N_0} (y - \boldsymbol{\alpha}^T \boldsymbol{\mu}_x). \quad (23)$$

Following (14)-(16), each \hat{x}_q can be modeled as $\hat{x}_q = \phi_q x_q + \xi_q$ where

$$\phi_q \equiv \frac{|\alpha_q|^2 \text{Var}(x_q)}{\sum_{k=0}^{Q-1} |\alpha_k|^2 \text{Var}(x_k) + N_0}. \quad (24)$$

It can also be verified that $\text{E}(\xi_q) = (1 - \phi_q)\text{E}(x_q)$ and $\text{Var}(\xi_q) = \phi_q(1 - \phi_q)\text{Var}(x_q)$. Then (17) becomes

$$\begin{aligned} v_q &\equiv \ln \left(\frac{\Pr(\hat{x}_q | x_q = +1) \Pr(x_q = +1)}{\Pr(\hat{x}_q | x_q = -1) \Pr(x_q = -1)} \right) \\ &= 2 \frac{\phi_q (\hat{x}_q - \text{E}(\xi_q))}{\text{Var}(\xi_q)} + \gamma_q. \end{aligned} \quad (25)$$

With some straightforward manipulations, it can be shown that (25) is equivalent to (9).

For a multi-path MIMO channel, the per symbol complexity with layer-by-layer detection is $O(Q + (1 + M/N) \log J + M^2)$. Interestingly, the normalized complexity per coded bit (i.e., per layer) does not increase with Q .

IV. IMPACT OF SIGNALING SCHEMES ON ITERATIVE LMMSE DETECTION

From the discussions in Section III-B, smaller $\text{Var}(x(i))$ leads to higher SNIR of the LMMSE estimate. In this section, we consider the minimization of $\text{Var}(x(i))$ in a statistical sense. We will investigate the impact of signaling schemes on the average value of $\text{Var}(x(i))$, which can be seen as the mean square error (MSE) in estimating $x(i)$ from the decoder feedback. We will prove that SCM can minimize this MSE during iterative LMMSE detection. This notable property of SCM indicates its potential performance advantage, which will be confirmed by the numerical results in Section V.

A. Gaussian Mapper

Denote by $\mathcal{R} : \mathcal{B} \rightarrow \mathcal{S}$ the mapping from a set of Q bits $B = \{b_0, b_1, \dots, b_{Q-1}\}$ to s_B in a constellation \mathcal{S} of size 2^Q with (1) as a special case. The signaling scheme is fully characterized by $(\mathcal{R}, \mathcal{S})$.

Denote by x an arbitrary entry in $\mathbf{x} = \{x(i)\}$. The block marked by ‘‘Gaussian mapper’’ in Fig. 3 produces the *a priori* mean and variance for x as follows. Let $\{\gamma_q\}$ be the *a priori* LLR values of $\{b_q\}$:

$$\gamma_q \equiv \ln \left(\frac{\Pr(b_q = 0)}{\Pr(b_q = 1)} \right), \quad q = 0, 1, \dots, Q-1. \quad (26)$$

During iterative detection, $\{\gamma_q\}$ are updated using the DEC feedback. From (26),

$$\Pr(b_q = 0) = 1 - \Pr(b_q = 1) = \frac{\exp(\gamma_q)}{1 + \exp(\gamma_q)}. \quad (27)$$

We now treat B as an integer with binary expression:

$$B = (b_0 b_1 \dots b_{Q-1}). \quad (28)$$

Then

$$\Pr(s_B) = \prod_{q=0}^{Q-1} \Pr(b_q), \quad (29)$$

where $\Pr(b_q)$ is either $\Pr(b_q = 0)$ or $\Pr(b_q = 1)$, depending on value of b_q . $\Pr(s_B)$ is the probability that the transmitted signal x is s_B . Therefore, we have

$$\text{E}(x) = \sum_{B=0}^{2^Q-1} s_B \Pr(s_B) \quad (30a)$$

$$\text{Var}(x) = \sum_{B=0}^{2^Q-1} |s_B - \text{E}(x)|^2 \Pr(s_B). \quad (30b)$$

The outputs of (30) are used to construct $\boldsymbol{\mu}_x$ and \mathbf{V}_x for the LMMSE estimation. (See (12).) $\text{Var}(x)$ in (30b) is a quadratic function of the constellation points $\{s_B\}$. This inspires us to find the optimum \mathcal{S} and \mathcal{R} that can minimize $\text{Var}(x)$ in a statistical sense, as detailed below.

B. Minimum MSE (MMSE)

The $\text{Var}(x)$ computed using (27)-(30) is clearly a function of $\{\gamma_q\}$. We now treat $\{\gamma_q\}$ as independent, identically distributed (i.i.d.) random variables drawn from a distribution $p_\gamma(\gamma)$ satisfying the symmetric condition [31] in (31) below.

Assumption I:

$$p_\gamma(\gamma) = p_\gamma(-\gamma). \quad (31)$$

We consider minimizing $\mathbb{E}[\text{Var}(x)]$, where $\mathbb{E}[\cdot]$ is the expectation over $p_\gamma(\gamma)$. Let $\rho \equiv \mathbb{E}[\text{Var}((-1)^{b_q})]$. (Note: ρ is not a function of q when $\{\gamma_q\}$ are i.i.d..)

Theorem I: Under constraint (3) and Assumption I and over all possible \mathcal{S} and \mathcal{R} ,

$$\text{MMSE} \equiv \min_{\mathcal{S}, \mathcal{R}} \mathbb{E}[\text{Var}(x)] = \rho. \quad (32)$$

Furthermore, SCM achieves the MMSE.

Proof: See Appendix.

The implications of the theorem are as follows. We can view $\mathbb{E}[\text{Var}(x)]$ as the MSE in estimating x using $\mathbb{E}(x)$ (where x is an arbitrary entry of \mathbf{x}). A smaller $\mathbb{E}[\text{Var}(x)]$ means that $\mathbb{E}(x)$ is statistically closer to the true value of x . Recall that the true *a priori* distribution of x is over \mathcal{S} , as specified by $\{\text{Pr}(s_B)\}$ in (29). Thus, the MSE in estimating x using $\mathbb{E}(x)$ depends on \mathcal{S} as well as $\mathcal{R} : \mathcal{B} \rightarrow \mathcal{S}$. Theorem I gives the minimum MSE, i.e., the MMSE, under the normalization constraint (3) and symmetry constraint (31). It also states that SCM can achieve this MMSE. This theorem together with the discussion in Section III-B shows that SCM is optimal for the Gaussian mapper.

Theorem I is illustrated by the numerical examples in Fig. 4. Following [31], we model $\{\gamma_q\}$ as independent samples from an AWGN channel, i.e., $\gamma \sim \mathcal{N}(2\mu d, 4\mu)$, $\forall \gamma \in \{\gamma_q\}$, where $d = \pm 1$ with equal probability and μ is the signal-to-noise ratio (SNR) of the channel. We characterize the reliability of $\{\gamma_q\}$ by the mutual information between $\{\gamma_q\}$ and $\{b_q\}$, denoted by I_γ . Fig. 4 compares the MSE versus I_γ curves for SCM and four other signaling schemes, namely, the 16-QAM with the modified set-partitioning (MSP), Mixed, Gray and M16^a mapping rules [11], [29]. The MMSE curve is generated using (32). Other curves are generated using Monte Carlo simulation for the signaling schemes compared. As expected, the SCM curve coincides with the MMSE one. Note that the 16-QAM with Gray mapping yields an MSE close to that for SCM. The MSE curves for other signaling methods are considerably poorer.

From Fig. 4, SCM is the optimal choice when the feedback variance is concerned. It is well known that other options such as MSP and M16^a have good Euclidean distance properties [11], [29]. The overall performance of the iterative LMMSE receiver depends on both these two factors, which is difficult to evaluate analytically. In the next section, we will assess the overall performance of iterative LMMSE detection based on simulation results.

V. NUMERICAL RESULTS

As mentioned earlier, the applications of the model in (10) include multi-user, MIMO and multi-path channels or their combinations. For the multi-user MIMO systems considered below, each user employs a BLAST-type scheme [35] (except that each coded bit sequence is randomly interleaved to facilitate iterative decoding) and mapped onto a constellation \mathcal{S} . The resultant symbol sequence is serial-to-parallel converted and then transmitted in parallel over different antennas.

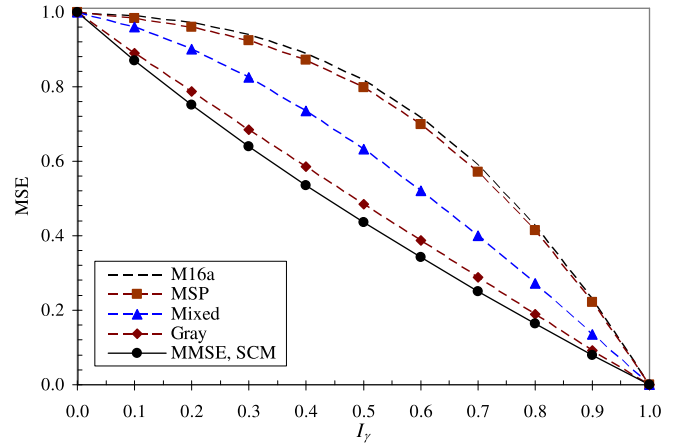


Fig. 4. Comparison of the MSE achieved by SCM (for SCM-1, SCM-2 and SCM-3) and four 16-QAM signaling schemes.

The SCM scheme can be formed by using a single code in the same way as conventional BICM-ID. An alternative approach is to employ multiple component codes following [14], [19]. We observed that the multi-code approach is advantageous if capacity-approaching component codes are used. However, the difference is not significant in most cases when relatively weak codes are used. For simplicity, we will adopt the single-code approach below except for the example with turbo codes in Fig. 5. BICM-Gray refers to the BICM-ID scheme based on Gray mapping. Similar notations apply to BICM-ID based on other mappings.

A. Performance in AWGN Channels with Turbo Codes

Fig. 5 shows examples based on the rate-1/2 turbo code $(23, 35)_8$ over AWGN channels. In this case, $\mathbf{x} = \{x(i)\}$ in (10) is a time-domain sequence and i is the time index. 16-ary signaling is used so the system rate is $R = 2$ bits/channel use. Multiple-code SCM using the constellation SCM-3 in Fig. 1(c) is compared with BICM-Gray and BICM-MSP using 16-QAM. For all these schemes, the APP de-mapping rule in (5) is directly applied; the number of (inner) iterations in the DEC is set to 6; and the number of (outer) iterations between the DEC and de-mapper (or ESE with SCM) is set to 4. The three schemes compared here have the same complexity.

It is well known that the Shannon limit for an AWGN channel is achieved by Gaussian signaling. In this example, we have only 16 constellation points and the irregular signaling scheme SCM-3 and the Gray signalling yield results close to the Shannon limit, as shown in Fig. 5. The MSP scheme with turbo code is not satisfactory. This is consistent with the observations in [38] that Gray mapping is advantageous over other alternatives in BICM-ID schemes with capacity-approaching codes.

B. Performance in More Complicated Channels

We now consider quasi-static MIMO channels. We assume that the channel coefficients are i.i.d. and quasi-static. We always assume that a rate-1/2, 4-state convolutional code $(5, 7)_8$ is used and the Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm [34] is applied to the DEC. For simplicity, we always use the single-code SCM schemes. We define E_b as the transmit energy per information bit.

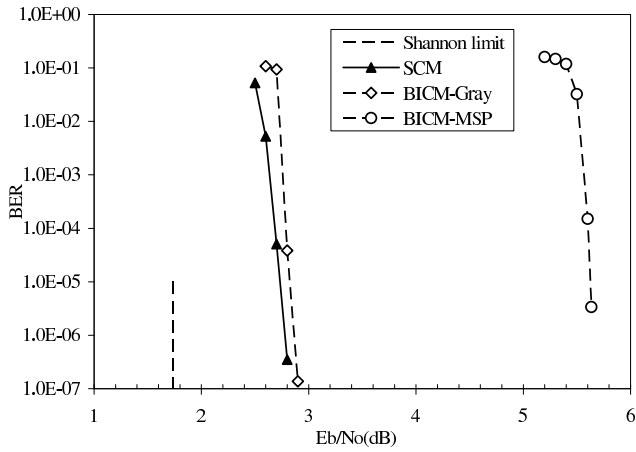


Fig. 5. Bit-error-rate (BER) of turbo-coded SCM and BICM-ID schemes over AWGN channels. The information block length is 65536. The system rate $R = 2$.

1) *Single-User Channel*: We first consider a single-user scheme with M receiver antennas and N transmitter antennas over single-path fading channel. In this case, an information block of 2048 bits is encoded into a binary sequence of length 4096 and then mapped into a 16-ary symbol sequence of length 1024. The symbol sequence is then partitioned into segments. Each segment is of the form

$$\mathbf{x} = \{x(i)\} = [x(1), x(2), \dots, x(N)]^T, \quad (33)$$

where $x(i)$ is transmitted from antenna i . The channel matrix is given by

$$\mathbf{H} = \begin{bmatrix} h(1,1) & h(1,2) & \dots & h(1,N) \\ h(2,1) & h(2,2) & \dots & h(2,N) \\ \vdots & \vdots & \ddots & \vdots \\ h(M,1) & h(M,2) & \dots & h(M,N) \end{bmatrix}, \quad (34)$$

where $h(m,n)$ is the channel coefficient between the n th transmit antenna and m th receive antenna. Let r be the rate of the encoder (ENC) in Fig. 2, K the number of users and N the number of transmit antennas per user. We define the system rate as

$$R = rQKN. \quad (35)$$

The LMMSE detection outlined in Section III-A can be directly applied. Fig. 6 compares the frame-error-rate (FER) with SCM and BICM-ID. The number of receiver antennas is fixed to $M = 4$. We consider $N = 1$ and 4 transmit antennas.

With $N = 1$, we apply MAP detection [33] for all the schemes compared. (The complexity of the MAP detection is still tolerable in this case.) The 16-QAM MSP signaling used in Fig. 6 is optimized for the MAP detection [33] and has a good distance profile. Indeed, we can see from Fig. 6 that BICM-MSP outperforms SCM in this case. Note that the advantage of SCM with respect to LMMSE estimation is not relevant here.

When $N = 4$, the MAP detection becomes too complicated, and so we resort to the LMMSE technique. In this case, SCM can significantly outperform BICM-MSP, which is mainly attributed to the MMSE property of SCM in Theorem I. Interestingly, when N increases from 1 to 4, the SCM

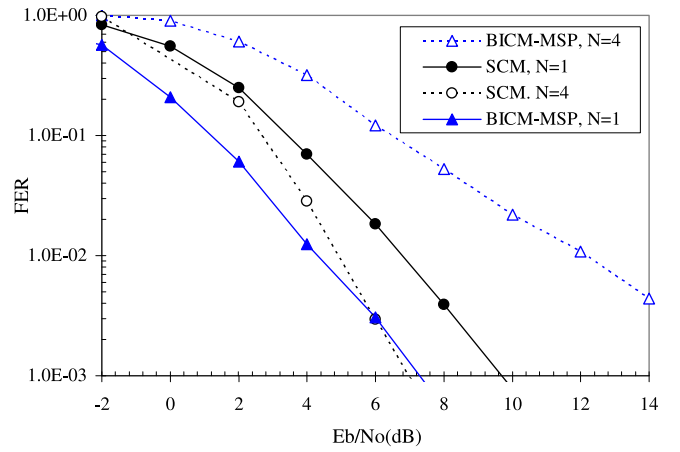


Fig. 6. FER of single-user MIMO systems with SCM and BICM-MSP over single-path channels. Different transmit antenna numbers N are considered. The receiver has $M = 4$ antennas. The constellation SCM-1 in Fig. 1 (a) is used. The information block length is 2048. The number of iterations is 10. The system rate $R = 2$ for $N = 1$ and $R = 8$ for $N = 4$.

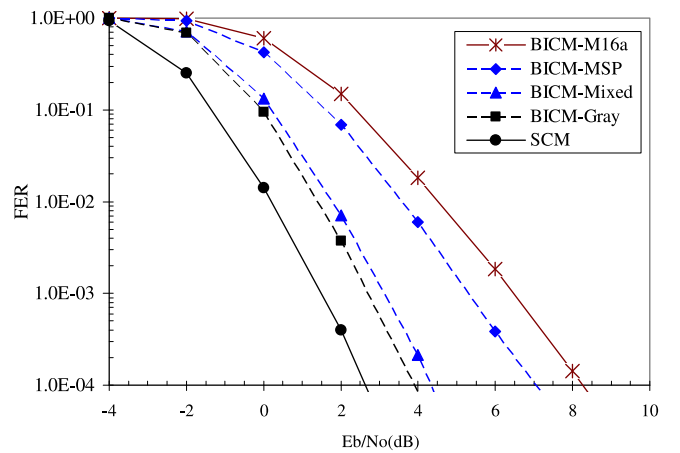


Fig. 7. FER for 2-user MIMO systems with SCM and BICM-ID over single-path channels. Each user has $N = 4$ transmit antennas and the receiver has $M = 8$ antennas. The constellation SCM-1 in Fig. 1(a) is used. The information block length is 1024 for each user. The number of iterations is 10. $R = 16$.

performance improves, thanks to the diversity gain offered by a larger N . In the meanwhile, the BICM-MSP performance degrades, indicating that iterative LMMSE estimation does not perform well with the MSP signaling.

2) *Multi-User MIMO Channel*: Next, we consider multiple users. In this case, define²

$$\mathbf{x} = \{x(i)\} = [(\mathbf{x}^{(0)})^T, (\mathbf{x}^{(1)})^T, \dots, (\mathbf{x}^{(K-1)})^T]^T, \quad (36)$$

$$\mathbf{H} = [\mathbf{H}^{(0)}, \mathbf{H}^{(1)}, \dots, \mathbf{H}^{(K-1)}], \quad (37)$$

where $\mathbf{x}^{(k)}$ and $\mathbf{H}^{(k)}$ denote the transmit signal segment and channel matrix for user- k , similar to those in (33) and (34). Now the index i in $\mathbf{x} = \{x(i)\}$ represents the combination of user and antenna index.

Fig. 7 presents the FER for various signaling schemes in a 2-

²We apply user-specific interleaving to $\{\mathbf{x}^{(k)}\}$ following the principles of the interleave-division multiple-access (IDMA) [36]. Note that, with \mathbf{H} and \mathbf{x} defined here, multi-user detection is implicitly realized in the iterative LMMSE process. See [24] for detail.

user case with the iterative LMMSE detector. The parameters are listed in the caption. We can see that the performance is roughly in line with the MSE comparison in Fig. 4, with SCM the best. Also note that BICM-Gray outperforms other BICM-ID schemes. Clearly, with LMMSE estimation, the performances of different signaling schemes are closely related to their MSE characteristics shown in Fig. 4.

3) *De-mapping Methods*: In the above examples, the APP de-mapping rule in (5) is applied after the LMMSE estimation. We now consider the “layer-by-layer” strategy mentioned in Section II-C for SCM that can greatly reduce de-mapping cost. Fig. 8 compares the performance of the two strategies in a single-user 2×2 MIMO system over a multi-path channel with $L = 4$ taps. The SCM signaling scheme is similar to that in Fig. 1(c), with parameters detailed in the caption. Layer-by-layer de-mapping results in certain performance degradation. Nevertheless, the achievable performance of SCM with layer-by-layer de-mapping is still better than that of BICM-Gray.

VI. CONCLUSIONS

We show that iterative LMMSE detection offers a low-cost option to high-rate coded systems over complicated channels. We conduct an MSE analysis and derive the MMSE solution for the Gaussian mapper involved in the LMMSE detection. We prove that SCM can maximize the SNIR during LMMSE detection and potentially outperform other alternatives. These claims are confirmed by simulation results for various multi-user/MIMO/multi-path channels. We have assumed that channel state information (CSI) is perfectly known in this paper. Discussions can be found in [39] for situations when CSI is not perfectly known but estimated.

APPENDIX

Let γ be the LLR of a bit $b \in \{b_q\}$. The following are easy to verify:

$$\Pr(b=0) = \frac{\exp(\gamma)}{1 + \exp(\gamma)}, \Pr(b=1) = \frac{1}{1 + \exp(\gamma)}, \quad (38a)$$

$$\mathbb{E}((-1)^b) = \Pr(b=0) - \Pr(b=1) = \frac{\exp(\gamma) - 1}{\exp(\gamma) + 1}, \quad (38b)$$

$$\begin{aligned} \text{Var}((-1)^b) &= 1 - \mathbb{E}((-1)^b)^2 = 1 - \left(\frac{\exp(\gamma) - 1}{\exp(\gamma) + 1}\right)^2 \\ &= 4\Pr(b=0)\Pr(b=1). \end{aligned} \quad (38c)$$

When γ is a random variable, the above quantities are also random variables. Assume that γ meets the symmetric condition (31). From (38),

$$\begin{aligned} \mathbb{E}[\Pr(b=0)] &= \mathbb{E}\left[\frac{\exp(\gamma)}{\exp(\gamma) + 1}\right] \\ &\stackrel{(a)}{=} \int_{-\infty}^{+\infty} \frac{\exp(\gamma)}{\exp(\gamma) + 1} p_{\gamma}(-\gamma) d\gamma \\ &\stackrel{\gamma' = -\gamma}{=} \int_{-\infty}^{+\infty} \frac{1}{\exp(\gamma') + 1} p_{\gamma}(\gamma') d\gamma' \\ &= \mathbb{E}[\Pr(b=1)], \end{aligned} \quad (39)$$

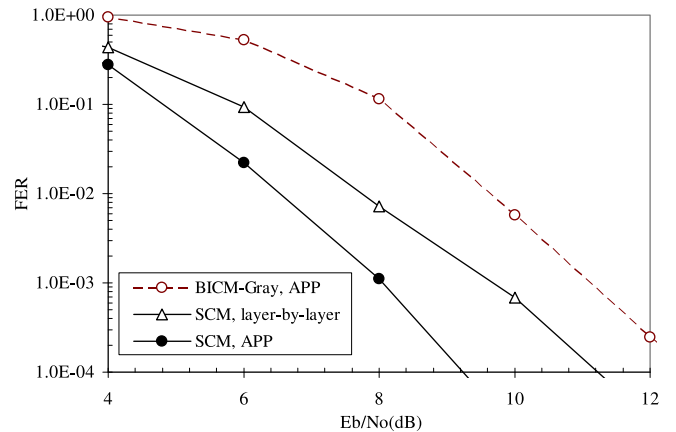


Fig. 8. Comparison of the APP and layer-by-layer de-mapping strategies when SCM is applied to a single-user MIMO system. For the SCM, $\{\alpha_q\} = \{\varepsilon, \varepsilon j, 1.5\varepsilon, 1.5\varepsilon j\}$, where $\varepsilon = \sqrt{2/13}$. $N = 2$, $M = 2$, $L = 4$ and $R = 4$. The information block length is 8192. The number of iterations is 20.

where equality (a) follows from (31). Since $\Pr(b=0) + \Pr(b=1) = 1$, (39) leads to

$$\mathbb{E}[\Pr(b=0)] = \mathbb{E}[\Pr(b=1)] = 1/2. \quad (40)$$

From the definition below (31), we have $\rho = \mathbb{E}[\text{Var}((-1)^b)]$. From (38c),

$$\rho = 4\mathbb{E}[\Pr(b=0)\Pr(b=1)]. \quad (41)$$

From $\Pr(b=0) + \Pr(b=1) = 1$, $\rho = 2 - 4\mathbb{E}[\Pr(b=0)^2] = 2 - 4\mathbb{E}[\Pr(b=1)^2]$. Thus,

$$\mathbb{E}[\Pr(b=0)^2] = \mathbb{E}[\Pr(b=1)^2] = \frac{1}{2} - \frac{\rho}{4}. \quad (42)$$

Denote by $\{s_B\}$ the 2^Q points in \mathcal{S} and define a vector $\mathbf{s} = [s_0, s_1, \dots, s_{2^Q-1}]^T$. Following (28), we treat $B = (b_0 b_1 \dots b_{Q-1})$ as a binary integer. From (30), we have

$$\mathbb{E}(x) = \sum_{x \in \mathcal{S}} x \Pr(x) = \mathbf{s}^T \mathbf{p}, \quad (43)$$

where \mathbf{p} is a vector formed by $\{\Pr(s_B)\}$. For example, when $Q = 2$, \mathcal{R} is

$$\begin{aligned} (b_0 = 0, b_1 = 0) &\rightarrow s_0, (b_0 = 0, b_1 = 1) \rightarrow s_1, \\ (b_0 = 1, b_1 = 0) &\rightarrow s_2, (b_0 = 1, b_1 = 1) \rightarrow s_3, \end{aligned}$$

and \mathbf{p} is given by (44) where “ \otimes ” denotes *Kronecker* product.

For a general Q , \mathbf{p} in (43) can be obtained using a chain of *Kronecker* products,

$$\mathbf{p} = \mathbf{p}_0 \otimes \dots \otimes \mathbf{p}_q \otimes \dots \otimes \mathbf{p}_{Q-1}, \quad (45)$$

where $\mathbf{p}_q = \begin{bmatrix} \Pr(b_q = 0) \\ \Pr(b_q = 1) \end{bmatrix}$. Define $\mathbf{Q}_q \equiv \mathbb{E}[\mathbf{p}_q \mathbf{p}_q^T]$, $\forall q$. From (42), we have the equation on the top of the following page and its eigenvalues $\{\lambda_0, \lambda_1\}$ and eigenvectors $\{\mathbf{g}_0, \mathbf{g}_1\}$ are

$$\begin{aligned} \lambda_0 &= 1/2 \text{ corresponding to } \mathbf{g}_0 = [1/2, 1/2]^T, \text{ and} \\ \lambda_1 &= (1 - \rho)/2 \text{ corresponding to } \mathbf{g}_1 = [1/2, -1/2]^T. \end{aligned}$$

$$\mathbf{p} = \begin{bmatrix} \Pr(b_0 = 0) \Pr(b_1 = 0) \\ \Pr(b_0 = 0) \Pr(b_1 = 1) \\ \Pr(b_0 = 1) \Pr(b_1 = 0) \\ \Pr(b_0 = 1) \Pr(b_1 = 1) \end{bmatrix} = \begin{bmatrix} \Pr(b_0 = 0) \\ \Pr(b_0 = 1) \end{bmatrix} \otimes \begin{bmatrix} \Pr(b_1 = 0) \\ \Pr(b_1 = 1) \end{bmatrix}. \quad (44)$$

$$\mathbf{Q}_q = \begin{bmatrix} \mathbb{E}[\Pr(b_q = 0)^2] & \mathbb{E}[\Pr(b_q = 1) \Pr(b_q = 0)] \\ \mathbb{E}[\Pr(b_q = 0) \Pr(b_q = 1)] & \mathbb{E}[\Pr(b_q = 1)^2] \end{bmatrix} \\ = \begin{bmatrix} 1/2 - \rho/4 & \rho/4 \\ \rho/4 & 1/2 - \rho/4 \end{bmatrix}$$

Define $\mathbf{Q} = \mathbb{E}[\mathbf{p}\mathbf{p}^T]$. From (44), we can see that

$$\mathbf{Q} = \mathbf{Q}_0 \otimes \cdots \otimes \mathbf{Q}_q \otimes \cdots \otimes \mathbf{Q}_{Q-1}. \quad (46)$$

From (46) and the spectrum property of *Kronecker* product [32], the eigenvalues of \mathbf{Q} are given by the diagonal of

$$\underbrace{\begin{bmatrix} 1/2 & 0 \\ 0 & (1-\rho)/2 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1/2 & 0 \\ 0 & (1-\rho)/2 \end{bmatrix}}_{Q \text{ times}}, \quad (47)$$

and the corresponding eigenvectors are given by the columns in

$$\underbrace{[\mathbf{g}_0, \mathbf{g}_1] \otimes \cdots \otimes [\mathbf{g}_0, \mathbf{g}_1]}_{Q \text{ times}}. \quad (48)$$

Since $0 \leq (\mathbb{E}((-1)^{b_q}))^2 \leq 1$, $0 \leq \rho = 1 - \mathbb{E}[(\mathbb{E}((-1)^{b_q}))^2] \leq 1$. Therefore, the largest eigenvalue in (47) is 2^{-Q} corresponding to eigenvector $2^{-Q/2}\mathbf{1}$, where $\mathbf{1}$ is an all-one vector with a proper size. Also, the second largest eigenvalue of \mathbf{Q} is $2^{-Q}(1-\rho)$ that corresponds to Q eigenvectors in (49) at the top of the next page. The vectors defined in (49) are normalized versions of the column vectors contained in

$$\mathbf{G} = \begin{bmatrix} +1 & \cdots & +1 & +1 \\ +1 & \cdots & +1 & -1 \\ +1 & \cdots & -1 & +1 \\ \vdots & & \vdots & \\ -1 & \cdots & -1 & -1 \end{bmatrix}, \quad (50)$$

where the B th row of \mathbf{G} forms the binary expression of B over $\{-1, +1\}$. From (40), we have

$$\mathbb{E}[\Pr(x)] = \prod_{q=0}^{Q-1} \mathbb{E}[\Pr(b_q)] = 2^{-Q}. \quad (51)$$

Then the MSE can be expressed as

$$\mathbb{E}[\text{Var}(x)] = \mathbb{E}\left[\sum_{x \in \mathcal{S}} |x|^2 \Pr(x) - |\mathbb{E}(x)|^2\right] \\ = \sum_{x \in \mathcal{S}} |x|^2 \mathbb{E}[\Pr(x)] - \mathbb{E}[|\mathbb{E}(x)|^2] \\ = 1 - \mathbf{s}^H \mathbf{Q} \mathbf{s}. \quad (52)$$

The last equality in (52) follows from (43) and (51). Minimizing (52) is equivalent to maximizing $\mathbf{s}^H \mathbf{Q} \mathbf{s}$, which is achieved when \mathbf{s} takes the direction of the eigenvector for the maximum eigenvalue of \mathbf{Q} . However, \mathbf{s} cannot be $2^{-Q/2}\mathbf{1}$ because it does not satisfy the condition $\sum_{x \in \mathcal{S}} x = 0$ in (3). (In fact,

$\mathbf{s} = 2^{-Q/2}\mathbf{1}$ implies that there is only one point in \mathcal{S} , and so it cannot be used to deliver information.) Furthermore, it can be verified from (48) that $\sum_{x \in \mathcal{S}} x = 0$ when \mathbf{s} takes the direction of any other eigenvector $\mathbf{g} \neq 2^{-Q/2}\mathbf{1}$. Thus, \mathbf{s} must be orthogonal to $2^{-Q/2}\mathbf{1}$ (otherwise $\sum_{x \in \mathcal{S}} x \neq 0$). Therefore we turn to the second largest eigenvalue $2^{-Q}(1-\rho)$. Then \mathbf{s} must fall in the space spanned by the columns of \mathbf{G} in (50), i.e.

$$\mathbf{s} = \mathbf{G} \boldsymbol{\alpha}, \quad (53)$$

for some $Q \times 1$ vector $\boldsymbol{\alpha} = \{\alpha_q\}$ with $\|\boldsymbol{\alpha}\|^2 = 1$. Thus

$$\text{MMSE} = \min_{\mathcal{S}, \mathcal{R}} \mathbb{E}[\text{Var}(x)] = \min_{\mathbf{s}} (1 - \mathbf{s}^H \mathbf{Q} \mathbf{s}) \\ = 1 - 2^Q \cdot 2^{-Q} \cdot (1 - \rho) = \rho. \quad (54)$$

Hence (32) holds. Finally, (53) is simply a vector form of (1) for the SCM with the constraints in (3).

REFERENCES

- [1] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 55–67, Jan. 1982.
- [2] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 371–377, May 1977.
- [3] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 5, pp. 873–884, May 1992.
- [4] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes," in *Proc. IEEE Int. Conf. Commun.*, Geneva, Switzerland, May 1993, pp. 1064–1070.
- [5] T. Richardson and R. Urbanke, "The capacity of low-density parity check codes under message-passing decoding," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.
- [6] S. L. Goff, A. Glavieux, and C. Berrou, "Turbo-codes and high spectral efficiency modulation," in *Proc. IEEE Int. Commun. Conf. (ICC'94)*, New Orleans, Mar. 1994, pp. 645–649.
- [7] P. Robertson and T. Worz, "Bandwidth-efficient turbo trellis-coded modulation using punctured component codes," *IEEE J. Select. Areas Commun.*, vol. 16, no. 2, pp. 206–218, Feb. 1998.
- [8] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Parallel concatenated trellis coded modulation," in *Proc. IEEE Int. Commun. Conf. (ICC'96)*, 1996, pp. 974–978.
- [9] D. J. Costello Jr., A. Banerjee, T. E. Fuja, and P. C. Massey, "Some reflections on the design of bandwidth efficient turbo codes," in *Proc. 4th Int. ITG Conf. Source & Channel Coding*, Berlin, Germany, Jan. 2002.
- [10] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inform. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.
- [11] A. Chindapol and J. A. Ritcey, "Design, analysis, and performance evaluation for BICM-ID with square QAM constellations in Rayleigh fading channels," *IEEE J. Select. Areas Commun.*, vol. 19, no. 5, pp. 944–957, May 2001.
- [12] T. M. Cover and J. M. Thomas, *Elements of Information Theory*, New York: Wiley, 1991.

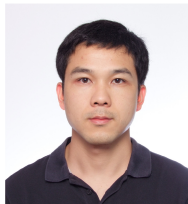
$$2^{-Q/2} \underbrace{\begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{q \text{ times}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \otimes \underbrace{\begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{Q-q-1 \text{ times}}, \quad 0 \leq q \leq Q-1 \quad (49)$$

- [13] L. Duan, B. Rimoldi, and R. Urbanke, "Approaching the AWGN channel capacity without active shaping," in *Proc. IEEE Int. Symp. Inform. Theory*, Ulm, Germany, June/July 1997, p. 374.
- [14] X. Ma and Li Ping, "Coded modulation using superimposed binary codes," *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 3331–3343, Dec. 2004.
- [15] P. A. Hoeher, H. Schoeneich, and J. Ch. Fricke, "Multi-layer interleaved multiple access: theory and practice," *European Trans. Telecommun.*, vol. 19, no. 5, pp. 523–536, Aug. 2008.
- [16] S. Gadkari and K. Rose, "Time-division versus superposition coded modulation schemes for unequal error protection," *IEEE Trans. Commun.*, vol. 47, no. 3, pp. 370–379, Mar. 1999.
- [17] N. H. Tran, H. H. Nguyen, and Tho. Le-Ngoc, "Performance of BICM-ID with signal space diversity," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1732–1742, May 2007.
- [18] T. W. Sun, R. D. Wesel, M. R. Shane, and K. Jarett, "Superposition turbo TCM for multirate broadcast," *IEEE Trans. Commun.*, vol. 52, no. 3, pp. 368–371, Mar. 2004.
- [19] J. Tong, Li Ping, and X. Ma, "Superposition coded modulation with peak-power limitation," *IEEE Trans. Inform. Theory*, vol. 55, no. 6, pp. 2562–2576, Jun. 2009.
- [20] M. Chen, T. Li, and O. M. Collins, "A computationally efficient multilevel coding scheme for ISI channels," *IEEE Trans. Inform. Theory*, vol. 53, no. 12, pp. 4556–4566, Dec. 2007.
- [21] X. Wang and H. V. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, pp. 1046–1061, July 1999.
- [22] M. Tuchler, R. Kowtner, and A. C. Singer, "Turbo equalization: principles and new results," *IEEE Trans. Commun.*, vol. 50, pp. 754–767, May 2002.
- [23] X. Wautlet, A. Dejonghe, and L. Vandendorpe, "MMSE-based fractional turbo receiver for space-time BICM over frequency-selective MIMO fading channels," *IEEE Trans. Signal Processing*, vol. 52, no. 6, pp. 1804–1809, June 2004.
- [24] X. Yuan, Q. Guo, and Li Ping, "Low-complexity iterative detection in multi-user MIMO ISI channels," *IEEE Signal. Process. Lett.*, vol. 15, pp. 25–28, 2008.
- [25] X. Yuan, Q. Guo, X. Wang, and Li Ping, "Evolution analysis of low-cost iterative equalization in coded linear systems with cyclic prefixes," *IEEE J. Select. Areas Commun.*, vol. 26, no. 2, pp. 301–310, Feb. 2008.
- [26] A. Dejonghe and L. Vandendorpe, "Bit-interleaved turbo equalization over static frequency-selective channels: Constellation mapping impact," *IEEE Trans. Commun.*, vol. 52, no. 12, pp. 2061–2065, Dec. 2004.
- [27] C. Hermosilla and L. Szczecinski, "Performance evaluation of linear turbo receivers using analytical extrinsic information transfer functions," *EURASIP J. Applied Signal Process.* 2005: 6, pp. 892–905
- [28] K. Wu and Li Ping, "A quasi random approach to space-time coding," *IEEE Trans. Inform. Theory*, vol. 54, no. 3, pp. 1073–1085, Mar. 2008.
- [29] F. Schreckenbach, N. Gortz, J. Hagenauer, and G. Bauch, "Optimization of symbol mappings for bit-interleaved coded modulation with iterative decoding," *IEEE Commun. Letters*, vol. 7, pp. 593–595, Dec. 2003.
- [30] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall PTR, 1993.
- [31] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.
- [32] S. R. Searle, *Matrix Algebra Useful for Statistics*, New York: John Wiley & Sons, 1982.
- [33] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York: Springer-Verlag, 1994.
- [34] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 284–287, Mar. 1974.
- [35] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, pp. 311–335, Mar. 1998.
- [36] Li Ping, L. H. Liu, K. Y. Wu, and W. K. Leung, "Interleave-division multiple-access," *IEEE Trans. Wireless Commun.*, vol. 5, no. 4, pp. 938–947, Apr. 2006.
- [37] Li Ping, W. K. Leung, and K. Y. Wu, "Low-rate turbo-Hadamard codes," *IEEE Trans. Inform. Theory*, vol. 49, no. 12, pp. 3213–3224, Dec. 2003.
- [38] J. Tan, "Iterative decoding and multicarrier modulation for wireless communications," Ph.D. dissertation, Georgia Inst. Technol., Atlanta, GA, Aug. 2003.
- [39] Q. Guo and Li Ping, "A low-complexity iterative channel estimation and detection technique for doubly selective channels," in *Proc. Globecom*, Dec. 2008.



Li Ping (S'87-M'91-SM'06) received his Ph.D. degree at Glasgow University in 1990. He lectured at Department of Electronic Engineering, Melbourne University, from 1990 to 1992, and worked as a member of research staff at Telecom Australia Research Laboratories from 1993 to 1995. He has been with the Department of Electronic Engineering, City University of Hong Kong, since January 1996, where he is now a chair professor. His research interests are communications systems and coding theory.

Dr. Li Ping received a British Telecom-Royal Society Fellowship in 1986, the IEE J J Thomson premium in 1993 and the Croucher Foundation award in 2005.



Jun Tong received the B.E. degree in information engineering and the M.E. degree in signal and information processing from the University of Electronic Science and Technology of China, Chengdu, in 2001 and 2004, respectively. He is currently working toward the Ph.D. degree at City University of Hong Kong. His research interests are signal processing and coding techniques in communication systems.



Xiaojun Yuan (S'04-M'08) received the B.S. degree in electronic and information systems from Shanghai Jiaotong University, the M.S. degree in circuit and systems from Fudan University, and the Ph.D. degree at City University of Hong Kong. He is currently working at University of Hawaii at Manoa, as a visiting research scholar. His research interests are signal processing, coding design, communication systems design and optimization, and networks.



Qinghua Guo (S'07-M'08) received the B.E. degree in electronic engineering and the M.E. degree in signal and information processing from Xidian University, Xi'an, China, in 2001 and 2004, respectively, and the Ph.D. degree in electronic engineering from City University of Hong Kong, Hong Kong SAR, in 2008. He is now with the School of Electrical, Electronic and Computer Engineering, The University of Western Australia, WA, Australia. His research interests include statistical signal processing and wireless communications.