

Impact of Signaling Schemes on Iterative Linear Minimum-Mean-Square-Error Detection

Li Ping, Jun Tong, Xiaojun Yuan, and Qinghua Guo
 Department of Electronic Engineering, City University of Hong Kong, Hong Kong
 E-mail: eeliping@cityu.edu.hk

Abstract—In this paper, we study the iterative detection problem for a coded system with multi-ary modulation. We show that, with iterative linear minimum-mean-square-error (LMMSE) detection, superposition coded modulation (SCM) can provide performance superior to that with other traditional signaling schemes used in trellis coded modulation (TCM) and bit-interleaved coded modulation (BICM). This finding provides a useful guideline for system design considering inter-symbol interference (ISI) and other forms of interference. Simulation results are provided to illustrate the efficiency of the iterative LMMSE detection with different signaling schemes.

I. INTRODUCTION

Consider the iterative detection problem for a system involving multi-ary modulation using a size- 2^M constellation. The channel may include, e.g., inter-symbol interference (ISI), multiple-access interference and cross-antenna interference. The complexity of the optimal receiver for such channels is usually prohibitively high, e.g., $O(2^{ML})$ for a detector involving L paths [1]. The iterative linear minimum-mean-square-error (LMMSE) detection provides a relatively low-cost alternative [2]–[6]. Good performance has been reported for such LMMSE receivers for binary phase shift keying (BPSK) [2], [3] or quadrature phase shift keying (QPSK) modulated signals [6].

It remains an interesting topic to examine the effectiveness of iterative LMMSE detection in systems involving multi-ary modulated signals. In this regard, it is reported in [7] that the quadrature-amplitude-modulation (QAM) with Gray mapping can outperform other options when iterative LMMSE detection is involved. It is shown [7] that the performance of an iterative LMMSE receiver is closely related to the signaling method at the transmitter side. This is because during iterative LMMSE detection, the accuracy of interference estimation is a function of signaling method. Such accuracy can be measured using the mean squared error (MSE) of interference estimation (for given feedbacks from the decoder).

In this paper, we establish the minimum limit for the MSE of interference estimation mentioned above. We show that this limit is achievable by superposition coded modulation (SCM) [8], [9]. The MSE achieved by QAM with Gray mapping is also quite close to this limit, but many other signaling schemes (e.g., those used for trellis coded modulation (TCM) and bit-interleaved coded modulation with iterative decoding (BICM-ID) [10]) are sub-optimal in this regard. Numerical results are provided to show that SCM can significantly outperform (in terms of bit-error-rate (BER) performance) other

alternative signaling schemes. We will also show that SCM is a good solution with respect to receiver complexity.

II. SYSTEM MODEL

A. Transmission Model

The transmitter scheme follows the principles of BICM-ID [10], as shown in the upper part of Fig. 1. The source data is first encoded by the encoder (ENC) using a binary forward-error-control (FEC) code, and permuted by an interleaver (marked by Π) to produce a bit sequence \mathbf{b} . Let \mathbf{b} be segmented into N sub-blocks

$$\mathbf{b} \equiv \{b(0), b(1), \dots, b(N-1)\} \quad (1)$$

where each $b(i)$ is a sub-block of M bits:

$$b(i) \equiv \{b^{(0)}(i), b^{(1)}(i), \dots, b^{(M-1)}(i)\}. \quad (2)$$

We naturally assume that each $b^{(m)}(i)$ is equally taken over $\{0, 1\}$. The mapper then maps each $b(i)$ onto a signaling point $x(i)$ in a constellation \mathcal{S} of size 2^M . The mapping rule $b(i) \rightarrow x(i) \in \mathcal{S}$ is denoted by \mathcal{R} .

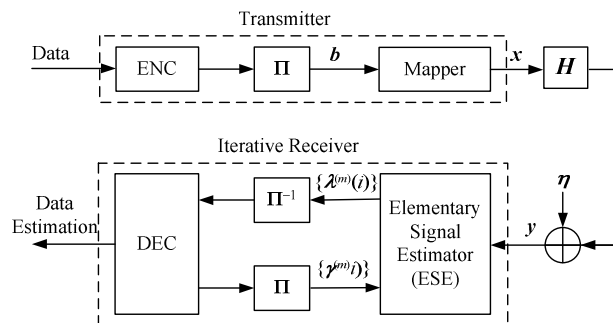


Fig. 1. The transmitter and iterative receiver structure of a coded multi-ary modulated system. Π denotes the interleaver and Π^{-1} the corresponding de-interleaver.

Let matrix \mathbf{H} represent the multiplicative effect of the channel. The received signal is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\eta}, \quad (3)$$

where \mathbf{y} is the received signal vector, \mathbf{x} the transmitted signal vector and $\boldsymbol{\eta}$ a vector of additive white Gaussian noise (AWGN) with mean $\mathbf{0}$ and covariance matrix $\sigma^2 \mathbf{I}$. In this paper, we always assume that \mathbf{H} is known perfectly at the receiver.

B. Iterative Detection Principles

The iterative receiver structure is shown in the lower part of Fig. 1. The elementary signal estimator (ESE) computes the *extrinsic* log-likelihood ratio (LLR) for each $b^{(m)}(i)$ as

$$\lambda^{(m)}(i) \equiv \ln \frac{\Pr(b^{(m)}(i) = 0 | \mathbf{y})}{\Pr(b^{(m)}(i) = 1 | \mathbf{y})} - \ln \frac{\Pr(b^{(m)}(i) = 0)}{\Pr(b^{(m)}(i) = 1)}, \quad \forall i, m, \quad (4)$$

with the FEC coding constraint ignored, i.e., the ESE operates as if \mathbf{b} contains un-coded bits. The decoder (DEC) performs *a posteriori* probability (APP) decoding using $\{\lambda^{(m)}(i)\}$ as the inputs, and producing the *extrinsic* LLRs

$$\gamma^{(m)}(i) \equiv \ln \frac{\Pr(b^{(m)}(i) = 0 | \{\lambda^{(m)}(i)\})}{\Pr(b^{(m)}(i) = 1 | \{\lambda^{(m)}(i)\})} - \ln \frac{\Pr(b^{(m)}(i) = 0)}{\Pr(b^{(m)}(i) = 1)}, \quad \forall i, m. \quad (5)$$

After decoding, the ESE operations can be executed again to refine the estimates in (4) using the feedbacks $\{\gamma^{(m)}(i)\}$. This process continues iteratively for a preset number of iterations. Hard decision is then performed in the final iteration to produce the data estimates. Detailed discussions on the above iterative detection process can be found in [2], [3], [5]. The APP decoding in (5) is a standard function. In what follows, we focus on the realization of the ESE function in (4).

C. The ESE Function

The following approach to the ESE is a low-cost, sub-optimal solution. As shown in Fig. 2, the detection process can be divided into the three steps listed below.

- (a) **Gaussian Approximation:** We approximate each $x(i)$ as a Gaussian random variable with mean $E(x(i))$ and variance $\text{Var}(x(i))$ computed using the DEC feedbacks $\{\gamma^{(m)}(i), \forall m\}$ (with details discussed in Section III). We assume that the entries of \mathbf{x} are uncorrelated, which can be (approximately) ensured using interleaving. We denote $E(\mathbf{x}) = [E(x(0)), E(x(1)), \dots, E(x(N-1))]^T$ and $\mathbf{V} \equiv \text{diag}(\text{Var}(x(0)), \text{Var}(x(1)), \dots, \text{Var}(x(N-1)))$.

- (b) **LMMSE Estimation:** Based on the Gaussian approximation, the LMMSE estimate of \mathbf{x} is [11]

$$\hat{\mathbf{x}} \equiv E(\mathbf{x} | \mathbf{y}) = E(\mathbf{x}) + \mathbf{V}\mathbf{H}^H \mathbf{R}^{-1}(\mathbf{y} - E(\mathbf{y})), \quad (6)$$

where $E(\mathbf{y}) \equiv \mathbf{H}E(\mathbf{x})$, and

$$\begin{aligned} \mathbf{R} &\equiv E((\mathbf{y} - E(\mathbf{y}))(\mathbf{y} - E(\mathbf{y}))^H) \\ &= \mathbf{H}\mathbf{V}\mathbf{H}^H + \sigma^2 \mathbf{I}. \end{aligned} \quad (7)$$

- (c) **Demapping:** We next calculate $\{\lambda^{(m)}(i), \forall m\}$ based on $\hat{x}(i)$, the i th entry of $\hat{\mathbf{x}}$. We rewrite $\hat{x}(i)$ as

$$\hat{x}(i) = \varphi(i)x(i) + \xi(i), \quad (8)$$

where $\varphi(i) \equiv \text{Var}(x(i))\mathbf{h}(i)^H \mathbf{R}^{-1} \mathbf{h}(i)$, $\mathbf{h}(i)$ is the i th column of \mathbf{H} , and $\xi(i)$ is assumed as a Gaussian noise independent of $x(i)$. Using (8), (4) can be implemented based on the maximum *a posteriori* probability (MAP) principle as

$$\begin{aligned} \lambda^{(m)}(i) &= \ln \frac{\sum_{s \in \mathcal{S}_m^0} \Pr(\hat{x}(i) | s) \Pr(s)}{\sum_{s \in \mathcal{S}_m^1} \Pr(\hat{x}(i) | s) \Pr(s)} - \gamma^{(m)}(i) \\ &= \ln \frac{\sum_{s \in \mathcal{S}_m^0} \exp\left(-\frac{|\hat{x}(i) - \varphi(i)s - E(\xi(i))|^2}{\text{Var}(\xi(i))}\right) \Pr(s)}{\sum_{s \in \mathcal{S}_m^1} \exp\left(-\frac{|\hat{x}(i) - \varphi(i)s - E(\xi(i))|^2}{\text{Var}(\xi(i))}\right) \Pr(s)} - \gamma^{(m)}(i) \end{aligned} \quad (9)$$

where \mathcal{S}_m^0 and \mathcal{S}_m^1 denote the subset of the constellation points in \mathcal{S} whose m th bit carriers 0 and 1, respectively.

In (9), $\Pr(s) = \prod_{m=0}^{M-1} \Pr(b^{(m)}(i))$ where $\Pr(b^{(m)}(i))$ can be computed from $\gamma^{(m)}(i)$ (as detailed in Section III). The complexity in (9) is $O(2^M)$.

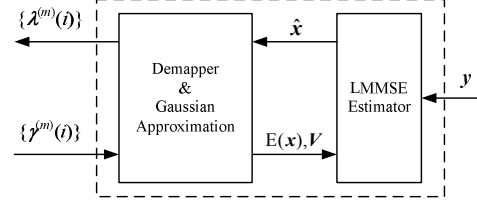


Fig. 2. The LMMSE approach to the ESE.

D. Discussions

Recall that \mathbf{x} is an N -dimensional vector with entries drawn from a constellation \mathcal{S} of size 2^M . The complexity for exactly evaluating (4) is $O(2^{MN})$ that is usually very high.

The discussion in Section II-C gives a low-cost alternative. Two approximations are involved here. First, each entry of \mathbf{x} is approximated by a continuous Gaussian variable in step (a) and, second, $\xi(i)$ is approximated by an additive Gaussian noise in step (c). With these two approximations, the complexity is reduced to $O(2^{M+N^2})$ (with $O(2^M)$ for step (a) and (c) and $O(N^2)$ for step (b)).

The impact of the first Gaussian approximation can be measured using $\text{Var}(x(i))$. A smaller $\text{Var}(x(i))$ implies that the first approximation is more accurate (as then $E(x(i))$ is statistically closer to the true value of $x(i)$). Interestingly, for given $\{\gamma^{(m)}(i)\}$, $\text{Var}(x(i))$ is a function of the signaling scheme, as we will see later. This implies that the accuracy of the first Gaussian approximation is different for different signaling methods. The choice of signaling methods also affects the second Gaussian approximation, since it can be shown that $\xi(i)$ in (8) is a function of $\{\text{Var}(x(i))\}$.

III. IMPACT OF SIGNALING SCHEMES

Continuing from Section II-D, we now consider minimizing $\{\text{Var}(x(i))\}$ in a statistical sense. For simplicity, we omit the time index i in this section unless it is necessary for discussion.

A. Signaling Scheme

Denote by $\mathcal{R}: b \rightarrow s$ the mapping from a set of M bits $b = \{b^{(0)}, b^{(1)}, \dots, b^{(M-1)}\}$ to a constellation point $s \in \mathcal{S}$ of size 2^M . We assume that \mathcal{S} is unbiased and with normalized power, i.e.,

$$\sum_{s \in \mathcal{S}} s = 0 \quad \text{and} \quad \frac{1}{2^M} \sum_{s \in \mathcal{S}} |s|^2 = 1. \quad (11)$$

The signaling scheme is then fully characterized by $(\mathcal{S}, \mathcal{R})$.

B. Estimation of Mean and Variance

Let $\{\gamma^{(m)}\}$ be the set of *a priori* LLR values of $\{b^{(m)}\}$ input to the ESE, i.e.,

$$\gamma^{(m)} \equiv \ln \frac{\Pr(b^{(m)}=0)}{\Pr(b^{(m)}=1)}, \quad m=0, 1, \dots, M-1. \quad (12)$$

In practice, $\{\gamma^{(m)}\}$ is updated using the feedbacks from the DEC. From (12),

$$\Pr(b^{(m)}=0) = 1 - \Pr(b^{(m)}=1) = \frac{\exp(\gamma^{(m)})}{1 + \exp(\gamma^{(m)})}. \quad (13)$$

For s mapped from a particular bit-combination $b = \{b^{(m)}\}$,

$$\Pr(s) = \prod_{m=0}^{M-1} \Pr(b^{(m)}), \quad (14)$$

where $\Pr(b^{(m)})$ is either $\Pr(b^{(m)}=0)$ or $\Pr(b^{(m)}=1)$, depending on mapping rule \mathcal{R} . Let x be the symbol associated with $\{\gamma^{(m)}\}$. Then, the mean and variance of x are, respectively,

$$E(x) = \sum_{s \in \mathcal{S}} s \Pr(s), \quad (15a)$$

$$\text{Var}(x) = \sum_{s \in \mathcal{S}} |s - E(x)|^2 \Pr(s). \quad (15b)$$

C. The MMSE in Gaussian Approximation

Clearly, $\text{Var}(x)$ in (15b) is a function of $\{\gamma^{(m)}\}$. We now treat $\{\gamma^{(m)}\}$ as random variables and consider minimizing $\mathbb{E}[\text{Var}(x)]$ where $\mathbb{E}[\cdot]$ is the expectation taken over the distribution of $\{\gamma^{(m)}\}$. Here $\mathbb{E}[\text{Var}(x)]$ can also be seen as the MSE in estimating x using $E(x)$. Note that $\mathbb{E}[\text{Var}(x)]$ is a function of \mathcal{S} and \mathcal{R} . The discussion below is to find the MMSE over all possible signaling methods with respect to the Gaussian approximation in Section II-C, which may potentially lead to improved performance.

Considering interleaving, we can treat $\{\gamma^{(m)}\}$ as i.i.d. random variables drawn from a distribution $p_\gamma(\gamma)$. Recall that $\{\gamma^{(m)}\}$ are updated using the feedback LLRs from the APP decoder. In this case, LLRs can be modeled as observations from an AWGN channel [3], [12]-[14] satisfying the following symmetric condition.

Assumption I: $p_\gamma(\gamma) = p_\gamma(-\gamma)$. (16)

Define $\rho \equiv \mathbb{E}[\text{Var}((-1)^{b^{(m)}})]$, $\forall m$. Here ρ is not a function of m since $\{\gamma^{(m)}\}$ are i.i.d.

Theorem I: Under Assumption I and over all possible \mathcal{S} satisfying (11) and mapping rules \mathcal{R} ,

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

Proof: See Appendix.

D. Superposition Coded Modulation (SCM)

SCM represents a special pair of \mathcal{S} and \mathcal{R} defined below.

Definition I: Given a set of M arbitrary complex coefficients $\{\alpha^{(m)}\}$ and given a binary bit set $b \equiv \{b^{(m)}\}$, the superposition mapping $\mathcal{R} : b \rightarrow s$ is defined as

$$s = \sum_{m=0}^{M-1} \alpha^{(m)} (-1)^{b^{(m)}}; \quad (17)$$

A superposition constellation \mathcal{S} is formed by running (17) over 2^M binary combinations of b .

Theorem II: The minimum $\mathbb{E}[\text{Var}(x)]$ given in Theorem I can be achieved by and only by SCM.

Proof: See Appendix.

Theorem I and II, together with the discussion in Section II-D, indicate that using SCM at the transmitter can potentially improve the performance of an iterative LMMSE detector. Some numerical examples are given later for illustration.

An additional advantage of SCM is its low complexity. Due to the similarity between the signalling in (17) and that of interleave-division multiple-access systems [6], the Gaussian-approximation-based detection method outlined in [6] can be applied to compute the demapper outputs for SCM. This approach has complexity $\mathcal{O}(M)$. For other conventional signalling schemes, the MAP method in (9) has to be used, which has complexity $\mathcal{O}(2^M)$.

E. Examples

From [14], $\{\gamma^{(m)}\}$ can be approximated as independent samples from an AWGN channel, i.e., $\gamma \sim \mathcal{N}(2\mu d, 4\mu)$, $\forall \gamma \in \{\gamma^{(m)}\}$, where $d = \pm 1$ with equal probability and μ is the signal-to-noise ratio (SNR) of the channel. Fig. 3 compares the MSE versus μ for SCM with that for three other signaling schemes, namely, the 16-QAM signaling with the modified set-partitioning (MSP), Mixed and Gray mappings [10]. SCM has a uniformly lower MSE than its alternatives, which agrees with Theorem I and II.

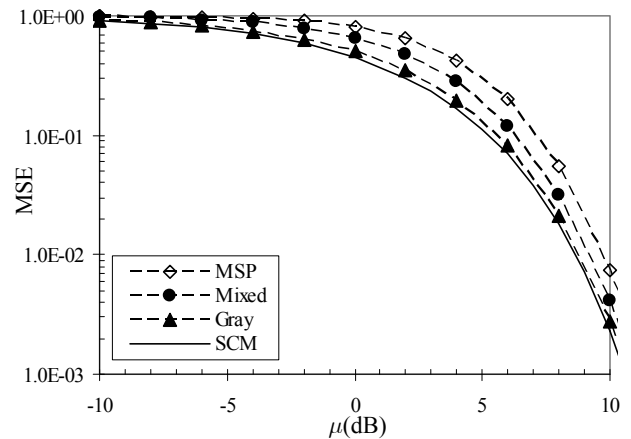


Fig. 3. Comparison of the MSE achieved by SCM and three other 16-QAM signaling schemes. For the SCM, $M = 4$, $\{\alpha^{(m)}\} = \{1, j, 1.5, 1.5j\}$, where $j = \sqrt{-1}$.

We next show the effect of minimized MSE on the system performance. Fig. 4 and Fig. 5 compare the BER performance of BICM-ID using the signaling schemes in Fig. 3 over the AWGN channel and the Porat channel [13], respectively. We focus on a target BER of 10^{-5} . In the AWGN channel, the LMMSE detector becomes the optimal MAP detector since there is no ISI in this case (i.e., $\xi(i)$ in (8) contains only channel noise). The minimization of MSE then has no effect on $\xi(i)$. In this case, as shown in Fig. 4, the 16-QAM with the Mixed and MSP mapping demonstrates better performance at $\text{BER} = 10^{-5}$. However, in the ISI channel where the LMMSE estimation is

necessary, the SCM performs better since it leads to improved performance of the LMMSE detector.

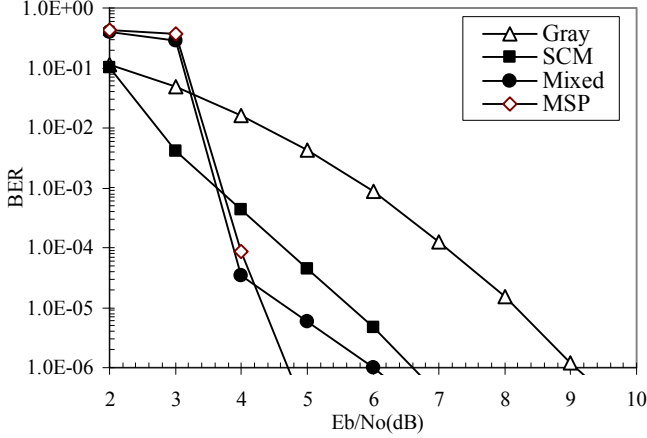


Fig. 4. Comparison of BICM-ID with different signaling schemes over the AWGN channel. A rate-1/2 convolutional code $(23, 35)_8$ is used and the information block length is 32768. System throughput = 2 bits/channel use.

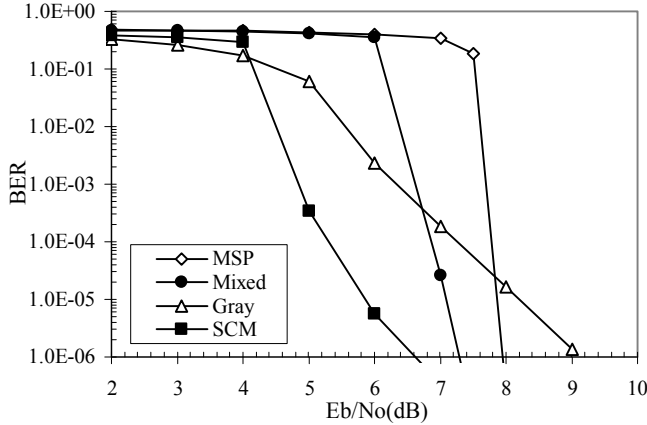


Fig. 5. Comparison of BICM-ID with different signaling schemes over the Porat channel. The coding parameters are the same as those in Fig. 4. The Porat channel [a 5-tap ISI channel $(2-0.4j, 1.5+1.8j, 1, 1.2-1.3j, 0.8+1.6j)$] is normalized in simulations.

V. CONCLUSIONS

We have shown by the MSE analysis that the SCM signaling is advantageous for iterative LMMSE detection. Numerical results demonstrate that SCM can outperform other conventional signaling schemes over single-user ISI channels. We are studying the applications of the finding of this paper in multiple-user and multiple-antenna systems.

APPENDIX: PROOF OF THEOREM I AND II

A. Preliminaries

In this subsection, we drop, with some abuse of notation, superscript m for $b^{(m)}$ and $\gamma^{(m)}$. Let γ be the LLR of bit b . The following are easy to verify:

$$\Pr(b=0) = \frac{\exp(\gamma)}{1+\exp(\gamma)} \text{ and } \Pr(b=1) = \frac{1}{1+\exp(\gamma)}, \quad (\text{A.1a})$$

$$\mathbb{E}[(-1)^b] = \Pr(b=0) - \Pr(b=1) = \frac{\exp(\gamma)-1}{\exp(\gamma)+1}, \quad (\text{A.1b})$$

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

When γ is a random variable, the above quantities are also random variables. Assume that γ meets the symmetric condition (16). From (A.1a) and (16),

$$\begin{aligned} \mathbb{E}[\Pr(b=0)] &= \mathbb{E} \left[\frac{\exp(\gamma)}{\exp(\gamma)+1} \right] \\ &\stackrel{(a)}{=} \int_{-\infty}^{+\infty} \frac{\exp(\gamma)}{\exp(\gamma)+1} p_{\gamma}(-\gamma) d\gamma \\ &= \int_{-\infty}^{+\infty} \frac{1}{\exp(\gamma')+1} p_{\gamma}(\gamma') d\gamma' \\ &= \mathbb{E}[\Pr(b=1)] \end{aligned} \quad (\text{A.2})$$

where equality (a) follows from (16). Since $\Pr(b=0) + \Pr(b=1) = 1$, (A.2) leads to

$$\mathbb{E}[\Pr(b=0)] = \mathbb{E}[\Pr(b=1)] = 1/2. \quad (\text{A.3})$$

From the definition below (16), we have $\rho = \mathbb{E}[\text{Var}((-1)^b)]$.

From (A.1c),

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

Again since $\Pr(b=0) + \Pr(b=1) = 1$,

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

Thus,

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

B. Proof of Theorem I

Now consider a constellation $S = \{s_b\}$ of size 2^M and the corresponding mapping rule $\mathcal{R}: b \rightarrow s \in S$. Define a vector $\mathbf{s} \equiv [s_0, s_1, \dots, s_{2^M-1}]^T$ where b (different from those in the previous subsection) is treated as an integer using binary expression,

$$b = (b^{(0)} \dots b^{(M-1)}).$$

From (15), we have

$$\mathbb{E}(x) = \sum_{s \in S} s \Pr(s) = \mathbf{s}^T \mathbf{p}, \quad (\text{A.5})$$

where \mathbf{p} is a vector formed by $\{\Pr(s_b)\}$. For example, when $M=2$, \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 ("⊗" for Kronecker product)

$$\mathbf{p} = \begin{bmatrix} \Pr(b^{(0)}=0) \Pr(b^{(1)}=0) \\ \Pr(b^{(0)}=0) \Pr(b^{(1)}=1) \\ \Pr(b^{(0)}=1) \Pr(b^{(1)}=0) \\ \Pr(b^{(0)}=1) \Pr(b^{(1)}=1) \end{bmatrix} = \begin{bmatrix} \Pr(b^{(0)}=0) \\ \Pr(b^{(0)}=1) \end{bmatrix} \otimes \begin{bmatrix} \Pr(b^{(1)}=0) \\ \Pr(b^{(1)}=1) \end{bmatrix}. \quad (\text{A.6})$$

For a general M , \mathbf{p} in (A.5) can be obtained using a chain of Kronecker products,

$$\mathbf{p} = \mathbf{p}^{(0)} \otimes \dots \otimes \mathbf{p}^{(m)} \otimes \dots \otimes \mathbf{p}^{(M-1)} \quad (\text{A.7})$$

where $\mathbf{p}^{(m)} = \begin{bmatrix} \Pr(b^{(m)}=0) \\ \Pr(b^{(m)}=1) \end{bmatrix}$. Define $\mathbf{Q}^{(m)} \equiv \mathbb{E}[\mathbf{p}^{(m)} \mathbf{p}^{(m)T}]$,

$\forall m$. From (A.4), we have

$$\mathbf{Q}^{(m)} = \begin{bmatrix} \mathbb{E}[\Pr(b^{(m)} = 0)^2] & \mathbb{E}[\Pr(b^{(m)} = 1)\Pr(b^{(m)} = 0)] \\ \mathbb{E}[\Pr(b^{(m)} = 0)\Pr(b^{(m)} = 1)] & \mathbb{E}[\Pr(b^{(m)} = 1)^2] \end{bmatrix}$$

$$= \begin{bmatrix} 1/2 - \rho/4 & \rho/4 \\ \rho/4 & 1/2 - \rho/4 \end{bmatrix},$$

and its eigenvalues $\{\lambda_0, \lambda_1\}$ and eigenvectors $\{\mathbf{g}_0, \mathbf{g}_1\}$ are $\lambda_0 = 1/2$ corresponding to $\mathbf{g}_0 = [1/2, 1/2]^T$, and $\lambda_1 = (1-\rho)/2$ corresponding to $\mathbf{g}_1 = [1/2, -1/2]^T$.

Define $\mathbf{Q} \equiv \mathbb{E}[\mathbf{p}\mathbf{p}^H]$. From (A.7), we can see that

$$\mathbf{Q} = \mathbf{Q}^{(0)} \otimes \dots \otimes \mathbf{Q}^{(m)} \otimes \dots \otimes \mathbf{Q}^{(M-1)}. \quad (\text{A.8})$$

From (A.8) and the spectrum property of *Kronecker* product [15], the eigenvalues of \mathbf{Q} is given by the diagonal of

$$\underbrace{\begin{bmatrix} 1/2 & 0 \\ 0 & (1-\rho)/2 \end{bmatrix} \otimes \dots \otimes \begin{bmatrix} 1/2 & 0 \\ 0 & (1-\rho)/2 \end{bmatrix}}_{M \text{ times}}, \quad (\text{A.9a})$$

and the corresponding eigenvectors given by the columns in

$$\underbrace{[\mathbf{g}_0, \mathbf{g}_1] \otimes \dots \otimes [\mathbf{g}_0, \mathbf{g}_1]}_{M \text{ times}}. \quad (\text{A.9b})$$

Since $0 \leq (\mathbb{E}[(-1)^{b^{(m)}}])^2 \leq 1$, thus $0 \leq \rho = 1 - \mathbb{E}[(\mathbb{E}[(-1)^{b^{(m)}}])^2] \leq 1$. Therefore, the largest eigenvalue in (A.9a) is 2^{-M} with corresponding eigenvector $2^{-M/2}\mathbf{1}$, where $\mathbf{1}$ is an all-one vector with proper size. Also, the second largest eigenvalue of \mathbf{Q} is $2^{-M}(1-\rho)$ that corresponds to M eigenvectors below

$$2^{-M/2} \underbrace{\begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes \dots \otimes \begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{m \text{ times}} \otimes \begin{bmatrix} 1 \\ -1 \end{bmatrix} \otimes \underbrace{\begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes \dots \otimes \begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{M-m-1 \text{ times}}, \quad 0 \leq m \leq M-1. \quad (\text{A.10a})$$

The vectors defined in (A.10a) are normalized versions of the column vectors contained in

$$\mathbf{G} = \begin{bmatrix} +1 & \dots & +1 & +1 \\ +1 & \dots & +1 & -1 \\ +1 & \dots & -1 & +1 \\ \vdots & & & \vdots \\ -1 & \dots & -1 & -1 \end{bmatrix}, \quad (\text{A.10b})$$

where the b th row of \mathbf{G} forms the binary expression of b over $\{-1, +1\}$. From (A.3), we have

$$\mathbb{E}[\Pr(s)] = \prod_{m=1}^M \mathbb{E}[\Pr(b^{(m)})] = 2^{-M}. \quad (\text{A.11})$$

Then,

$$\begin{aligned} \mathbb{E}[\text{Var}(x)] &= \mathbb{E}\left[\sum_{s \in \mathcal{S}} |s|^2 \Pr(s) - |\mathbb{E}(x)|^2\right] \\ &= \sum_{s \in \mathcal{S}} |s|^2 \mathbb{E}[\Pr(s)] - \mathbb{E}[|\mathbb{E}(x)|^2] \\ &= 1 - \mathbf{s}^H \mathbf{Q} \mathbf{s}, \end{aligned} \quad (\text{A.12})$$

The last equality in (A.12) follows from (A.5) and (A.11). Minimizing (A.12) 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 $2^{-M/2}\mathbf{1}$ because it does not satisfy the condition $\sum_{s \in \mathcal{S}} s = 0$ in (11). Furthermore, it can be verified from (A.9b) that $\sum_{s \in \mathcal{S}} s = 0$ when \mathbf{s} takes the direction of any other eigenvector $\mathbf{g} \neq 2^{-M/2}\mathbf{1}$. Thus, \mathbf{s} must be orthogonal to $2^{-M/2}\mathbf{1}$ (otherwise $\sum_{s \in \mathcal{S}} s \neq 0$).

Therefore we turn to the second largest eigenvalue $2^{-M}(1-\rho)$. Then \mathbf{s} must fall in the space spanned by the columns of \mathbf{G} in (A.10b), i.e.,

$$\mathbf{s} = \mathbf{G} \boldsymbol{\alpha}, \quad (\text{A.13})$$

for any $M \times 1$ vector $\boldsymbol{\alpha}$ with $\|\boldsymbol{\alpha}\|^2 = 1$. Thus,

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

C. Proof of Theorem II

Eqn. (A.13) is simply a vector form expression of (17) for the SCM with the constraints in (11).

ACKNOWLEDGMENT

This work was fully supported by a grant from the Research Grant Council of the Hong Kong SAR, China [Project No. CityU 117305].

REFERENCES

- [1] S. Verdú, *Multi-user Detection*, Cambridge University Press, 1998.
- [2] X. Wang and H. V. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1046–1061, July 1999.
- [3] M. Tüchler, R. Koetter, and A. C. Singer, "Turbo equalization: principles and new results," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 754–767, May 2002.
- [4] D. Bokolamulla and T. Aulin, "Multiuser detection for continuous phase modulation over Rayleigh fading channels," *IEEE Commun. Letters*, vol. 9, no. 10, pp. 906–908, Oct. 2005.
- [5] A. Dejonghe and L. Vandendorpe, "Turbo-equalization for multilevel modulation: an efficient low-complexity scheme," in *Proc. IEEE Int. Conf. Communications (ICC)*, New York, USA, May 2002, pp. 1863–1867.
- [6] Li Ping, L. Liu, K. Wu, and W. K. Leung, "Interleave division multiple-access," *IEEE Trans. Wireless Commun.*, vol. 5, no. 4, pp. 938–947, Apr. 2006.
- [7] C. Hermosilla and L. Szczecinski, "Performance evaluation of linear turbo receivers using analytical extrinsic information transfer functions," *EURASIP J. Applied Signal Process.*, vol. 2005, no. 6, pp. 892–905, May 2005.
- [8] L. Duan, B. Rimoldi, and R. Urbanke, "Approaching the AWGN channel capacity without active shaping," in *Proc. ISIT'97*, Ulm, Germany, p. 374, July 1997.
- [9] X. Ma and Li Ping, "Coded modulation using superimposed binary codes," *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 3331–3343, Dec. 2004.
- [10] 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.
- [11] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, 1993.
- [12] X. Yuan, Q. Guo, 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.
- [13] V. Ramon, C. Herzet, and L. Vandendorpe, "A semi-analytical method for predicting the performance and convergence behavior of a multiuser Turbo-equalizer/demapper," *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 1104–1117, Mar. 2007.
- [14] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.
- [15] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, 1991.