Achievable Rates of MIMO Systems With Linear Precoding and Iterative LMMSE Detection

Xiaojun Yuan, *Member, IEEE*, Li Ping, *Fellow, IEEE*, Chongbin Xu, and Aleksandar Kavcic, *Senior Member, IEEE*

***Abstract*— We establish area theorems for iterative detection and decoding (or simply, iterative detection) over coded linear systems, including multiple-input multiple-output channels, inter-symbol interference channels, and orthogonal frequency-division multiplexing systems. We propose a linear precoding technique that asymptotically ensures the Gaussianness of the messages passed in iterative detection, as the transmission block length tends to infinity. Area theorems are established to characterize the behavior of the iterative receiver. We show that, for unconstrained signaling, the proposed single-code scheme with linear precoding and iterative linear minimum mean-square error (LMMSE) detection is potentially information lossless, under various assumptions on the availability of the channel state information at the transmitter. We further show that, for constrained signaling, our proposed single-code scheme considerably outperforms the conventional multicode parallel transmission scheme based on singular value decomposition and water-filling power allocation. Numerical results are provided to verify our analysis.**

***Index Terms*— Linear precoding, iterative LMMSE detection, area theorem, superposition coded modulation.**

I. Introduction

A. Area Properties

THE optimal detection of a coded signal in a complicated wireless environment may incur excessive computational complexity. Iterative detection provides a low-cost solution by decomposing the overall receiver into two or more local processors and conducting message-passing between the local processors to refine the detection output (see [1]–[11] and the references therein). The analysis of an iterative detection process is an intriguing problem. The density-evolution technique [2] shows that carefully designed low-density parity-check (LDPC) codes can achieve near-capacity performance in additive white Gaussian noise (AWGN) channels with iterative message-passing decoding algorithms. It was further shown in [12] that the achievable rate of an iterative scheme for an erasure channel can be measured by the area under the so-called extrinsic information transfer (EXIT) curves [13] and the channel capacity is approachable when the two local processors have matched EXIT curves. This area property is extended in [14] to scalar AWGN channels (or simply, AWGN channels) using the measure of minimum mean-square error (MMSE), in which a sufficient condition is established to approach the capacity of a binary-input AWGN channel with iterative detection.

It is commonly accepted that, with random interleaving, the extrinsic information (i.e., the messages) from an *a posteriori* probability (APP) decoder for a binary forward-error-control code can be modeled as a sequence of observations from an effective scalar AWGN channel [13]. Thus, the authors in [14] made two basic assumptions: (i) the messages passed between the local processors are modeled as the observations from an effective scalar AWGN channel (referred to as the AWGN assumption) and (ii) the local processors are optimal in the sense of APP detection/decoding. Assumption (i) ensures that each local processor can be characterized by a single-variable transfer function involving signal-to-noise ratio (SNR) and MMSE. The mutual information and MMSE relationship established in [15] is then applied to derive the area property, and capacity-approaching performance is proven when the transfer curves of the local processors are matched.

B. MIMO Channels

The work in [14] is limited to scalar channels. A straightforward extension to multiple-input multiple-output (MIMO) channels is to convert a MIMO channel into a set of independent scalar sub-channels using linear processing based on singular value decomposition (SVD), to allocate power among these sub-channels using the water-filling (WF) principle [44] or its variation mercury water-filling [16], and then to apply variable-rate coding to these sub-channels (with one codebook assigned to one sub-channel). Hence the results in [14] can be applied to each sub-channel individually. This multi-code SVD-WF approach with variable-rate coding is conceptually simple, but may encounter the following difficulties in practice.

First, the rate and power pair obtained by SVD-WF for each sub-channel varies and so a large number of encoders with various rates are required. The related complexity can be

Manuscript received March 18, 2012; revised August 6, 2014; accepted August 6, 2014. Date of publication August 19, 2014; date of current version October 16, 2014. This work was supported by the Research Grant Council of Hong Kong under Grant 418712 and Grant CityU 118013. This paper was presented at the 2011 International Symposium on Information Theory.

X. Yuan is with the School of Information Science and Technology, ShanghaiTech University, Shanghai 200031, China (e-mail: yuanxj@shanghaitech.edu.cn).

L. Ping and C. Xu are with the Department of Electrical Engineering, City University of Hong Kong, Hong Kong (e-mail: eeliping@cityu.edu.hk; xchongbin2@cityu.edu.hk).

A. Kavcic is with the Department of Electrical Engineering, University of Hawaii at Manoa, Honolulu, HI 96822 USA (e-mail: kavcic@hawaii.edu).

Communicated by D. Guo, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2014.2349513

reduced by rate quantization, but performance may deteriorate noticeably when quantization interval increases.

Second, this SVD-WF approach requires accurate channel state information at the transmitter (CSIT). In practice, CSIT error can be caused by the Doppler effect or limited backhaul capacity. In general, perfect orthogonality between the sub-channels cannot be established in the presence of CSIT error. Then the optimal transceiver design based on the SVD-WF principle becomes a complicated issue.

The above difficulties motivate us to study alternative transmission techniques for MIMO systems without orthogonal channel decomposition. Linear precoding (LP) [21]–[26] can be used for this purpose. Iterative detection can be applied to MIMO systems with LP. However, the extension of the area property in [14] to MIMO systems without orthogonal channel decomposition is not straightforward. This is because a multivariate function is required in such a system to characterize the receiver's behavior, but then assumption (i) mentioned earlier may not hold. Also, optimal decoding for a MIMO system (without orthogonal channel decomposition) can be too costly, especially in the recent development of massive MIMO systems. Linear MMSE (LMMSE) detection [18] has the advantage of low complexity, but it is not optimal for non-Gaussian signaling, implying that assumption (ii) may not hold either.

C. Contributions of This Paper

In this paper, we consider a joint LP and iterative LMMSE detection scheme. We show that the proposed linear precoding technique ensures that the AWGN assumption asymptotically holds for the output of the LMMSE detector, provided that the transmission block length is sufficiently large. This allows us to use a single pair of input and output parameters to characterize the behavior of the LMMSE detector. Also, we adopt superposition coded modulation (SCM) [19], [20] to approximate Gaussian signaling, for which the local LMMSE detection is near-optimal. These treatments ensure the validity of the two assumptions in Section I.A for MIMO setups. Based on this, we establish area theorems for the proposed LP and iterative LMMSE detection (LP-LMMSE) scheme in MIMO channels, under the assumptions of perfect and partial CSIT, respectively. We show that, for unconstrained signaling, the proposed LP-LMMSE scheme with a single code is potentially information lossless based on the curve-matching principle, even though the suboptimal iterative LMMSE detection technique is employed.

The above area theorems provide guidelines in the practical design of the LP-LMMSE scheme. Specifically, we show that optimal power allocation for non-Gaussian inputs can be obtained by solving a convex optimization problem. Also, efficient practical FEC codes, such as LDPC codes, can be designed based on the curve-matching principle. Furthermore, we show that, for non-Gaussian signaling, the proposed single-code LP-LMMSE scheme achieves much higher rates than the multi-code SVD-WF scheme when the encoding and decoding complexities of the two schemes are kept at a comparable level.

From information theory, the capacity of a MIMO channel can be achieved using a single random code with SVD and WF [58]. However, the realization of such a single-code scheme with affordable encoding and decoding complexity remains a challenging problem. As such, one of the major contributions of this paper is to show a practical single-code scheme with LP and iterative LMMSE detection that achieves the MIMO capacity.

D. Comparisons With Existing LP Techniques

It is interesting to compare the proposed linear precoding technique with other alternatives (see [21]–[28] for full CSIT and [29]–[34] for partial CSIT). From an information-theoretic viewpoint, the ultimate criterion for precoder design is to achieve the channel capacity (see [23], [33], [34] and the references therein). It is well known that channel coding is required to achieve the capacity. However, most existing works on precoder design focus on un-coded MIMO systems equipped with a signal detector at the receiver performing far from optimal.

Along this line, a variety of design criteria have been studied, such as pair-wise error probability minimization [22], max-min signal-to-interference-plus-noise ratio (SINR) [23], average SINR maximization [23], SINR equalization [24], [25], and MMSE [26], [27], etc. The works for un-coded systems can be applied to coded systems by concatenating the detector with a decoder. Unfortunately, the residual interference at the output of the detector in general results in a considerable performance loss [36], [55], [56].

Iterative detection and decoding can efficiently suppress the residual interference left by the detector. Then, a major challenge is to jointly design FEC coding and linear precoding at the transmitter, taking into account the effect of iterative detection at the receiver. This paper provides a simple solution to this problem. Our analysis shows that, with curve-matching codes, the proposed LP-LMMSE scheme is capacity-achieving under various assumptions on CSIT.

E. Comparisons With the SVD-WF Approach

As aforementioned, the conventional SVD-WF approach is conceptually simple, provided that ideal Gaussian signaling is used and perfect CSIT is available. For non-Gaussian inputs, the mercury water-filling (MWF) technique [16] can be used for power allocation among different eigen-modes after SVD. However, it is shown in [41] and [42] that the optimal linear precoder for non-Gaussian inputs is in general non-diagonal even for independent parallel channels. To the best of our knowledge, it is computationally intensive to determine the optimal precoder in this case.

Furthermore, the SVD-WF approach requires perfect CSIT that is difficult to acquire in practice, e.g., due to the Doppler effect or limited backhaul capacity. CSIT uncertainty in general impairs the orthogonality between the parallel sub-channels. Thus, efficient interference cancellation techniques, such as iterative detection and decoding, are required at the receiver [4], [5], [7], [10].

The above discussions imply that linear precoding and iterative detection are necessary in the practical design of an efficient MIMO transceiver. In this regard, the LP-LMMSE scheme developed in this paper provides a low-cost solution to the problem. In particular, only one encoder is required in the proposed LP-LMMSE scheme. This is much simpler than the SVD-WF scheme that requires multiple encoders with various rates. Further, the proposed LP-LMMSE scheme can achieve higher information rates than SVD-WF and its variation SVD-MWF for practical systems with non-Gaussian inputs, as demonstrated by the numerical results provided later in Section VI.

F. Outline of the Remainder of the Paper

Section II describes our system model and iterative LMMSE detection under consideration. Section III describes linear precoding for MIMO systems with perfect CSIT, and characterizes the corresponding iterative receiver using SINR-variance evolution. In Section IV, we establish the area theorems and consider the precoder design. Section V extends our results to MIMO channels with CSIT uncertainty. Section VI provides numerical results, and Section VII offers some concluding remarks.

II. System Model and Iterative LMMSE Detection

A. Generic Linear System

A generic complex-valued linear system is modeled as

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\eta} \tag{1}$$

where $\boldsymbol{y}$ is a J'-by-1 received signal vector, $\boldsymbol{A}$ is a J'-by-J transfer matrix, $\boldsymbol{x} = [x_1, x_2, \ldots, x_J]^{\mathrm{T}}$ is a J-by-1 transmit signal vector, and $\boldsymbol{\eta} \sim \mathcal{CN}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$ a J'-by-1 additive noise vector. Notations $\boldsymbol{0}$ and $\boldsymbol{I}$, respectively, represent an all-zero vector and an identity matrix with a proper size. Note that both J' and J are integers, with the meanings revealed later.

Throughout this paper, we assume full channel state information at the receiver, i.e., the receiver perfectly knows the transfer matrix $\boldsymbol{A}$. We will discuss the situations of both perfect and partial CSIT.

B. Messages

The transmitter structure for the system in (1) is shown in the upper part of Fig. 1. The forward-error-correction (FEC) encoder generates a frame of JK coded symbols, denoted by

$$\boldsymbol{x}' = [x'_1, x'_2, \ldots, x'_{JK}]^{\mathrm{T}}$$

where K is the number of the uses of system (1) in a codeword frame. These coded symbols are randomly interleaved by the interleaver Π and then partitioned into K segments with equal length J. Each segment serves as an input $\boldsymbol{x}$ to the system (1). We emphasize that $\boldsymbol{x}$ and $\boldsymbol{x}'$ are not of the same length, for that $\boldsymbol{x}'$ represents the symbols in an overall codeword frame consisting of K segments, while $\boldsymbol{x}$ only represents the symbols in (any) one of these segments.

The iterative receiver, as illustrated in the lower half of Fig. 1, consists of two local processors, namely, the detector and the decoder, inter-connected by the interleaver Π and the de-interleaver Π^{-1}. Particularly, the detector considered in this paper follows the LMMSE principle, hence the name *LMMSE estimator*.

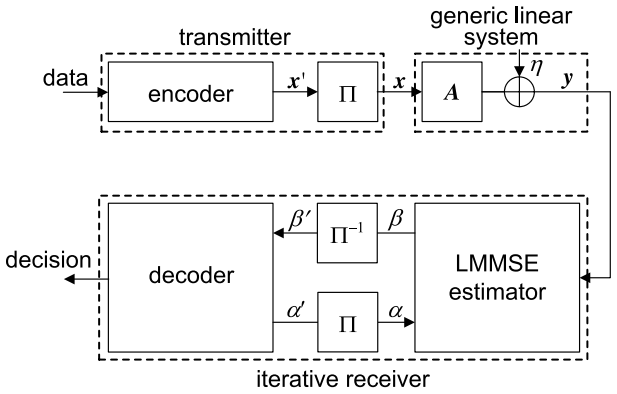

Fig. 1. The transceiver structure of the proposed scheme over the generic linear system in (1). Π is the interleaver and Π^{-1} is the corresponding de-interleaver.

The LMMSE estimator estimates $\boldsymbol{x}$ based on the channel observation $\boldsymbol{y}$ and the messages from the decoder (denoted by α). The outputs of the estimator are extrinsic messages (denoted by β). The message set α is defined as follows. Denote by x_i the ith entry of $\boldsymbol{x}$. Each x_i is constrained on a discrete signaling constellation $\mathcal{S} = \{s_1, s_2, \ldots, s_{|\mathcal{S}|}\}$, where $|\mathcal{S}|$ represents the size of $\mathcal{S}$. We always assume that each x_i is randomly and uniformly taken over the constellation points in $\mathcal{S}$. This assumption does not lose any generality since signal shaping (e.g., to approach Gaussian signaling) can be realized by properly designing the constellation points of $\mathcal{S}$.

For each use of system (1), let α be a vector of J messages as

$$\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_J]^{\mathrm{T}},$$

and each message α_i is a set of $|\mathcal{S}|$ likelihood values for x_i, i.e.

$$\alpha_i = \{\alpha_i(1), \alpha_i(2), \ldots, \alpha_i(|\mathcal{S}|)\},$$

where $\alpha_i(k)$ represents the likelihood of $x_i = s_k \in \mathcal{S}$ prior to the processing of the detector, and

$$\sum_{k=1}^{|\mathcal{S}|} \alpha_i(k) = 1.$$

Similar notation applies to the message vector $\beta = [\beta_1, \beta_2, \ldots, \beta_J]^{\mathrm{T}}$.

The decoder decodes $\boldsymbol{x}'$ based on the input message set β', and outputs the extrinsic message set α'. Decoding is performed on the overall frame, and so both α' and β' contain JK messages. Let x'_i be the ith entry of $\boldsymbol{x}'$, and $\beta'_i(k)$ be the likelihood of $x'_i = s_k \in \mathcal{S}$ prior to decoding. Then, we can express

$$\beta' = [\beta'_1, \beta'_2, \ldots, \beta'_{JK}]^{\mathrm{T}}$$

where $\beta'_i = \{\beta'_i(1), \beta'_i(2), \ldots, \beta'_i(|\mathcal{S}|)\}$ for $i = 1, \ldots, JK$. Messages in α' allow similar expressions. After decoding, α' are interleaved and partitioned to form the input of the LMMSE estimator, which completes one round of iteration.

The LMMSE estimator and the decoder are executed iteratively until convergence.

C. Basic Assumptions

Here we discuss some basic assumptions used throughout this paper. For notational convenience, we introduce an auxiliary random variable, denoted by a_i, to represent the information carried by each α_i. The conditional probability of a_i given x_i is defined as

$$p(a_i|x_i = s_k) = \alpha_i(k).$$

Similarly, an auxiliary random variable b'_i is defined for each β'_i.[1] The conditional probability of b'_i given x'_i is given by

$$p(b'_i|x'_i = s_k) = \beta'_i(k).$$

Denote $\boldsymbol{a} = [a_1, \ldots, a_J]^{\mathrm{T}}$ and $\boldsymbol{b}' = [b'_1, \ldots, b'_{JK}]^{\mathrm{T}}$. We make the following assumptions on the inputs of the local processors.

Assumption 1: For the detector, each x_i is independently drawn from $\mathcal{S}$ with equal probability, i.e., $p(x_i = s_k) = 1/|\mathcal{S}|$, for any i and k; the messages in $\boldsymbol{a}$ are conditionally independent and identically distributed (i.i.d.) given $\boldsymbol{x}$, i.e.

$$p(\boldsymbol{a}|\boldsymbol{x}) = \prod_{i=1}^{J} p(a_i|x_i),$$

and for any dummy variable t and $k = 1, 2, \ldots, |\mathcal{S}|$,

$$p(a_i = t|x_i = s_k) = p(a_j = t|x_j = s_k).$$

Assumption 2: For the decoder, the messages in $\boldsymbol{b}'$ are conditionally i.i.d. given $\boldsymbol{x}'$, i.e.

$$p(\boldsymbol{b}'|\boldsymbol{x}') = \prod_{i=1}^{JK} p(b'_i|x'_i),$$

and for any dummy variable t and $k = 1, 2, \ldots, |\mathcal{S}|$,

$$p(b'_i = t|x'_i = s_k) = p(b'_j = t|x'_j = s_k).$$

With Assumption 1, the joint probability space of $\boldsymbol{x}$, $\boldsymbol{a}$, and $\boldsymbol{y}$ seen by the detector can be expressed as

$$p(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{y}) = \left(\prod_{i=1}^{J} p(x_i) p(a_i|x_i) \right) p(\boldsymbol{y}|\boldsymbol{x}), \tag{2a}$$

where $p(\boldsymbol{y}|\boldsymbol{x})$ is determined by (1). Similarly, with Assumption 2, the joint probability space of $\boldsymbol{x}'$ and $\boldsymbol{b}'$ seen by the decoder is

$$p(\boldsymbol{x}', \boldsymbol{b}') = p(\boldsymbol{x}') \prod_{i=1}^{JK} p(b'_i|x'_i) \tag{2b}$$

where $\boldsymbol{x}'$ is assumed to be uniformly taken over the codebook, i.e., $p(\boldsymbol{x}')$ is a uniform distribution over the codebook.

Remark 1: Assumptions 1 and 2 decouple the probability space seen by the detector and decoder, which simplifies the analysis of the iterative process. Similar assumptions have been widely used in iterative decoding and turbo equalization algorithms [2], [13], [17], [37]. We emphasize that, in fact, Assumptions 1 and 2 are required only for those involved symbols in the computation tree of the iterative algorithm at a depth of the maximum number of iterations. Assumption 1 only involves J symbols, and thus can be made valid by random interleaving as K tends to infinity. However, strictly speaking, Assumption 2 cannot be justified in this way, for that this assumption involves JK symbols and the corresponding messages are generally correlated (due to the LMMSE detection operation), no matter what interleaver is used. A remedy to this issue is to require Assumption 2 hold only for the neighborhood of the computation tree, instead of for all the JK symbols in the codeword. This remedy will make the derivation of the results in this paper more rigorous, but will complicate the notation and the related discussions. For ease of presentation, we do not adopt this remedy and keep Assumption 2 as it is.

D. LMMSE Estimation

The detector delivers the *extrinsic* messages defined as

$$\beta_i(k) = p\left(x_i = s_k | \boldsymbol{a}_{\sim i}, \boldsymbol{y}\right), \text{ for } i = 1, 2, \ldots, J \text{ and } k = 1, \ldots, |\mathcal{S}|$$

where the joint probability space of $\boldsymbol{x}$, $\boldsymbol{a}$, and $\boldsymbol{y}$ is given in (2a), and $\boldsymbol{a}_{\sim i}$ represents the vector obtained by deleting the ith entry of $\boldsymbol{a}$. Here, "extrinsic" means that the contribution of the *a priori* message a_i is excluded in calculating each β_i.

For system (1), a direct evaluation of $\beta_i(k)$ may be excessively complicated. The LMMSE estimation is a low cost alternative. It is suboptimal in general, but it is optimal if $\boldsymbol{x}$ are generated using Gaussian signaling.

Denote the mean and covariance of $\boldsymbol{x}$ (seen by the detector) by $\bar{\boldsymbol{x}} = [\bar{x}_1, \ldots, \bar{x}_J]^{\mathrm{T}}$ and $v\boldsymbol{I}$, respectively, with

$$\bar{x}_i = \mathrm{E}[x_i|a_i] = \sum_{k=1}^{|\mathcal{S}|} \alpha_i(k) s_k \tag{3a}$$

and

$$v = \mathrm{E}\left[|x_i - \bar{x}_i|^2 \right], \quad \text{for any index } i, \tag{3b}$$

where the expectation is taken over the joint distribution of a_i and x_i. From Assumption 1, v is invariant with respect to the index i. In practice, v is approximated by the sample variance as

$$v \approx J^{-1} \sum_{i=1}^{J} \sum_{k=1}^{|\mathcal{S}|} \alpha_i(k) \, |s_k - \bar{x}_i|^2,$$

which converges to the true variance when $J \to \infty$.

The LMMSE estimator of $\boldsymbol{x}$ given $\boldsymbol{y}$ is [38]

$$\hat{\boldsymbol{x}} = \bar{\boldsymbol{x}} + v\boldsymbol{A}^{\mathrm{H}} \boldsymbol{R}^{-1} (\boldsymbol{y} - \boldsymbol{A}\bar{\boldsymbol{x}}) \tag{4a}$$

where $\boldsymbol{R}$ is the covariance matrix of $\boldsymbol{y}$ given as

$$\boldsymbol{R} = v\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}} + \sigma^2 \boldsymbol{I}. \tag{4b}$$

Recall that the extrinsic message β_i should be independent of a_i. To meet this requirement, we calculate the extrinsic mean and variance for x_i (denoted by b_i and u_i, respectively)

[1] The auxiliary random variables are introduced for notational convenience. Usually, an iterative receiver directly evaluates β' without explicitly calculating $\{b'_i\}$. However, we will see later that $\{b'_i\}$ are indeed calculated in the iterative LMMSE detection, based on which β' is determined.

by excluding the contribution of a_i according to the Gaussian message combining rule (cf., (54) and (55) in [39]) as

$$u_i^{-1} = M(i,i)^{-1} - v^{-1} \tag{5a}$$

and

$$\frac{b_i}{u_i} = \frac{\hat{x}_i}{M(i,i)} - \frac{\bar{x}_i}{v}, \tag{5b}$$

where $M(i,i)$ represents the (i,i)th entry of the MMSE matrix

$$\boldsymbol{M} = v\boldsymbol{I} - v^2\boldsymbol{A}^{\mathrm{H}}\boldsymbol{R}^{-1}\boldsymbol{A}. \tag{5c}$$

Finally, the output messages of the detector can be calculated as

$$\beta_i(k) = \frac{p(b_i|x_i = s_k)}{\sum_{k=1}^{|\mathcal{S}|} p(b_i|x_i = s_k)}, \quad \text{for any } i \text{ and } k. \tag{5d}$$

In the above, $\beta_i(k)$ can be readily calculated by assuming that b_i is an observation of x_i over an effective AWGN channel with noise power u_i. The justification of this assumption can be found in Lemma 1 in Section III.C. We note that b_i in (5b) is just the auxiliary variable of $\beta_i(k)$ defined in Section II.C; see Footnote 1.

E. APP Decoding

The decoder decodes $\boldsymbol{x}'$ based on the messages β' following the APP decoding principle. The extrinsic output of the decoder for each x_i' is defined as

$$\alpha_i'(k) = p\left(x_i' = s_k \middle| \boldsymbol{b}'_{\sim i}\right) \tag{6}$$

for $i = 1, 2, \ldots, JK$ and $k = 1, \ldots, |\mathcal{S}|$, where the joint probability space of $\boldsymbol{x}'$ and $\boldsymbol{b}'$ is given in (2b).

Assumption 3: The local decoder performs APP decoding.

In practice, APP decoding is usually computationally expensive. Low-complexity message-passing algorithms can be used to achieve near-optimal performance, provided that the code structure allows a sparse graphic description [37]. Message-passing decoding is well-studied in the literature, and thus the details are omitted here. Compared with APP decoding, the performance loss of message-passing decoding is usually marginal. This loss is not of concern in this paper. Therefore, we introduce Assumption 3 to simplify our analysis.

III. Linear Precoding With Perfect CSIT

In this section, we assume perfect CSIT. The case of imperfect CSIT will be discussed in Section V. We propose a linear precoding technique to equalize the output SINRs of the detector. We then establish an SINR-variance transfer chart technique to analyze the performance of the proposed LP-LMMSE scheme.

A. MIMO Channels

The iterative LMMSE detection principle described in the previous section can be applied to any coded linear systems, including MIMO channels, inter-symbol interference (ISI) channels, and orthogonal frequency-division multiplexing (OFDM) systems, etc. In the following, as an example, we establish a connection between a MIMO channel and the system in (1).

A Gaussian MIMO channel with N transmit and M receive antennas can be modeled as

$$\tilde{\boldsymbol{y}}_l = \boldsymbol{H}\tilde{\boldsymbol{x}}_l + \tilde{\boldsymbol{\eta}}_l \tag{7}$$

where l represents the l-th channel use, $\tilde{\boldsymbol{y}}_l$ is an M-by-1 received signal vector, $\boldsymbol{H}$ is the M-by-N channel transfer matrix known at both the transmitter and the receiver, $\tilde{\boldsymbol{x}}_l$ is an N-by-1 transmit signal vector, and $\tilde{\boldsymbol{\eta}}_l \sim \mathcal{CN}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$ is an M-by-1 additive noise vector.

A transmission block (i.e., one use of system (1)) involves J/N uses of the channel (7), where J is the block length defined in Section II.A. We assume that J' and J are properly chosen so that $J'/M = J/N$ is an integer (representing the number of channel uses in system (1)). Denote

$$\begin{aligned}\tilde{\boldsymbol{y}} &= [\tilde{\boldsymbol{y}}_1^{\mathrm{T}}, \ldots, \tilde{\boldsymbol{y}}_{J/N}^{\mathrm{T}}]^{\mathrm{T}}\\ \tilde{\boldsymbol{x}} &= [\tilde{\boldsymbol{x}}_1^{\mathrm{T}}, \ldots, \tilde{\boldsymbol{x}}_{J/N}^{\mathrm{T}}]^{\mathrm{T}}\\ \tilde{\boldsymbol{\eta}} &= [\tilde{\boldsymbol{\eta}}_1^{\mathrm{T}}, \ldots, \tilde{\boldsymbol{\eta}}_{J/N}^{\mathrm{T}}]^{\mathrm{T}}.\end{aligned}$$

Combining J/N channel uses, we write an extended system as

$$\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{H}}\tilde{\boldsymbol{x}} + \tilde{\boldsymbol{\eta}} \tag{8a}$$

where the extended channel is given by

$$\tilde{\boldsymbol{H}} = \boldsymbol{I}_{J/N} \otimes \boldsymbol{H} \tag{8b}$$

with "$\otimes$" being the Kronecker product and $\boldsymbol{I}_n$ being an n-by-n identity matrix. The signal power is constrained as

$$\mathrm{E}\left[\tilde{\boldsymbol{x}}^{\mathrm{H}}\tilde{\boldsymbol{x}}\right]/J \le P. \tag{8c}$$

The SVD of $\tilde{\boldsymbol{H}}$ is given by

$$\tilde{\boldsymbol{H}} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{H}} \tag{9}$$

where $\boldsymbol{\Lambda}$ is an (JM/N)-by-J diagonal matrix with non-negative diagonal elements, and $\boldsymbol{U}$ and $\boldsymbol{V}$ are unitary matrices. We assume that $\boldsymbol{U}$ and $\boldsymbol{V}$ are chosen such that the diagonal entries of $\boldsymbol{\Lambda}$ are asymptotically uncorrelated as J tends to infinity.[2] This property is useful in establishing Lemma 1 in Section III.C.

B. Linear Precoding

We focus on the following linear precoding operation:

$$\tilde{\boldsymbol{x}} = \boldsymbol{V}\boldsymbol{W}^{1/2}\boldsymbol{F}\boldsymbol{x} \tag{10a}$$

where $\tilde{\boldsymbol{x}}$ and $\boldsymbol{x}$ are given in (8a) and (1), respectively, $\boldsymbol{V}$ is defined in (9), $\boldsymbol{W}$ is a diagonal matrix with non-negative diagonal elements for power allocation, and $\boldsymbol{F}$ is the normalized DFT matrix with the (i,k)th entry given by

$$F(i,k) = J^{-1/2}\exp(-\mathrm{j}2\pi(i-1)(k-1)/J) \tag{10b}$$

with $\mathrm{j} = \sqrt{-1}$.

[2]Although $\tilde{\boldsymbol{H}}$ is deterministic, the ordering of its singular values can be arbitrarily chosen. Here we choose such a *random* ordering that the diagonal of $\boldsymbol{\Lambda}$ is an asymptotically uncorrelated sequence as J tends to infinity.

From Assumption 1, the entries of $\boldsymbol{x}$ are uncorrelated. Without loss of generality, we further assume that the entries of $\boldsymbol{x}$ have normalized power. Denote by $W(i,\, i)$ the $(i,\, i)$-th diagonal entry of $\boldsymbol{W}$. Then the power constraint in (8c) becomes

$$J^{-1}\sum_{i=1}^{J} W(i,i) \le P. \tag{11}$$

The optimization of $\boldsymbol{W}$ will be detailed in Section IV.

The use of $\boldsymbol{F}$ in (10a) is to ensure that the SINRs are equal for all symbols after LMMSE detection; see Section III.C. Incidentally, the choice of the DFT matrix for $\boldsymbol{F}$ also allows the fast Fourier transform (FFT) algorithm in the implementation of LMMSE detection [10]. We note that similar DFT-based precoders have been used in MIMO systems for other purposes, e.g., for harnessing diversity in [46].

At the receiver side, the received vector $\tilde{\boldsymbol{y}}$ is post-processed by the matrix $\boldsymbol{U}^{\mathrm{H}}$:

$$\boldsymbol{y} = \boldsymbol{U}^{\mathrm{H}}\tilde{\boldsymbol{y}}. \tag{12}$$

Combining (8)–(12) and letting

$$\boldsymbol{D} = \boldsymbol{\Lambda}\boldsymbol{W}^{1/2}, \tag{13a}$$

we obtain an equivalent channel as

$$\boldsymbol{y} = \boldsymbol{D}\boldsymbol{F}\boldsymbol{x} + \boldsymbol{U}^{\mathrm{H}}\tilde{\boldsymbol{\eta}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\eta} \tag{13b}$$

where $\boldsymbol{A} = \boldsymbol{D}\boldsymbol{F}$ and $\boldsymbol{\eta} = \boldsymbol{U}^{\mathrm{H}}\tilde{\boldsymbol{\eta}} \sim \mathcal{CN}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$ since $\boldsymbol{U}$ is unitary. Clearly, (13b) has the same form as (1). The iterative detection and decoding procedure outlined in Section II can be directly applied to (13b).

C. Characterization of the Estimator

We now aim to characterize the behavior of the estimator. It was shown in [23] that, with the precoder in (10a), the SINR becomes uniform for (non-iterative) LMMSE estimation. We next show that a similar property holds in our proposed iterative system. We further show that, for the extended system in (8a), the residual interference in the output of the estimator is asymptotically Gaussian, following the central limit theorem.

To start with, we establish the fact that $M(i,i)$, the ith diagonal element of $\boldsymbol{M}$ in (5c), is not a function of i. To see this, we substitute (4b) and $\boldsymbol{A} = \boldsymbol{D}\boldsymbol{F}$ into (5c):

$$\boldsymbol{M} = v\boldsymbol{I} - v^2\boldsymbol{F}^{\mathrm{H}}\boldsymbol{D}^{\mathrm{H}}(v\boldsymbol{D}\boldsymbol{D}^{\mathrm{H}} + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{D}\boldsymbol{F}.$$

Then

$$M(i,i) = v - v^2\boldsymbol{f}_i^{\mathrm{H}}\boldsymbol{D}^{\mathrm{H}}(v\boldsymbol{D}\boldsymbol{D}^{\mathrm{H}} + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{D}\boldsymbol{f}_i \tag{14a}$$

$$= v - J^{-1}\sum_{k=1}^{J}\frac{v^2 D(k,k)^2}{vD(k,k)^2 + \sigma^2} \tag{14b}$$

$$= \frac{1}{J}\sum_{k=1}^{J}\left(v^{-1} + \frac{D(k,k)^2}{\sigma^2}\right)^{-1} \tag{14c}$$

where $\boldsymbol{f}_i$ is the ith column of $\boldsymbol{F}$, and (14b) utilizes (10b). Clearly, $M(i,i)$ is the same for all i.

We now introduce a definition:

$$\phi(v) = \frac{1}{M(i,i)} - \frac{1}{v}.$$

Note that $\phi(v)$ defined above is not a function of the index i, as $M(i,i)$ is invariant with respect to i. Recall that $\{D(k,k) = \Lambda(k,k)W(k,k)^{1/2}, k = 1, 2, \ldots, J\}$ and that $\{\Lambda(k,k)\}$ are the singular values of $\tilde{\boldsymbol{H}}$ in (8b) and so $\{\Lambda(k,k)\}$ contains J/N copies of the singular values of $\boldsymbol{H}$ (denoted by $\lambda_1, \ldots, \lambda_N$). We assume that a same amount of power is allocated to any two eigen-modes with the same channel gains, i.e., if $\Lambda(k,k) = \Lambda(j,j)$, then $W(k,k) = W(j,j)$.[3] Then, $\{D(k,k)\}$ have at most N different values, denoted by $d_1, \ldots, d_N$. Together with (14c) and the definition of $\phi(v)$ above, we obtain

$$\phi(v) = \left(\frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{v} + \frac{d_n^2}{\sigma^2}\right)^{-1}\right)^{-1} - v^{-1}. \tag{15}$$

Remark 2: We describe some useful properties of $\phi(v)$. First, as the signal power (i.e., the power of each entry of $\boldsymbol{x}$) is normalized to 1, the variance v in (15) varies from 0 to 1. Second, $\phi(v)$ is monotonically decreasing in v, implying that for the LMMSE estimator the higher the input variance, the lower the output SINR. Third, as $v \to 0, \phi(v)$ tends to the average SNR given by

$$\phi(0) = \frac{1}{N}\sum_{n=1}^{N}\frac{d_n^2}{\sigma^2}.$$

Remark 3: We emphasize that $\boldsymbol{F}$ is chosen to meet the following criteria: (a) $\boldsymbol{F}$ is unitary; (b) $\boldsymbol{F}\boldsymbol{x}$ allows fast computation with complexity $O(J\log J)$; (c) the diagonal elements of $\boldsymbol{F}\boldsymbol{T}\boldsymbol{F}^{\mathrm{H}}$ is invariant for any real diagonal matrix $\boldsymbol{T}$ (and so are the diagonal elements of $\boldsymbol{M}$). The choice of $\boldsymbol{F}$ to meet these criteria is not unique. Besides the DFT matrix, we may alternatively choose $\boldsymbol{F}$ to be, e.g., the Hadamard matrix.

We next show that the outputs of the LMMSE estimator can be modeled as the observations from an AWGN channel provided that J is sufficiently large, and the related SINR is given by $\rho = \phi(v)$. Define

$$n_i \triangleq \frac{v}{M(i,i)\rho}\boldsymbol{f}_i^{\mathrm{H}}\boldsymbol{D}^{\mathrm{H}}(v\,\boldsymbol{D}\boldsymbol{D}^{\mathrm{H}} + \sigma^2\boldsymbol{I})^{-1}(\boldsymbol{A}(\boldsymbol{x}_{\setminus i} - \bar{\boldsymbol{x}}_{\setminus i}) + \boldsymbol{\eta}) \tag{16}$$

where $\boldsymbol{x}_{\setminus \mathrm{i}}$ (or $\bar{\boldsymbol{x}}_{\setminus i}$) represents the vector obtained by setting the ith entry of $\boldsymbol{x}$ (or $\bar{\boldsymbol{x}}$) to zero. It is clear that n_i is independent of x_i. Moreover, we have the following result with the proof given in Appendix A.

Lemma 1: For any index i, b_i in (5d) can be expressed as

$$b_i = x_i + n_i, \tag{17}$$

where n_i is independent of x_i and its distribution converges to $\mathcal{CN}(0,\, 1/\phi(v))$ as $J \to \infty$.

In the above, n_i represents the residual interference plus noise at the output of the LMMSE estimator in iteration. Hence, $\rho = \phi(v)$ in (15) represents the related SINR. It is well-known that the residue error of the LMMSE estimation is approximately Gaussian [38], [52]. Lemma 1 reveals that

[3] This treatment doesn't incur any information loss, as seen from Theorem 1.

this approximation becomes exact for the precoder in (10a) over the extended channel in (8a) for a sufficiently large J.

We henceforth always assume that J is sufficiently large, so that each b_i can be modeled as an observation of x_i from an effective AWGN channel with SNR $= \phi(v)$.

D. Characterization of the Decoder

We now consider the characterization of the decoder's behavior. The decoder performs APP decoding upon receiving the messages $\boldsymbol{b}'$ modeled as independent observations of $\boldsymbol{x}'$ over an AWGN channel with SNR $= \rho$ (cf., Lemma 1). Define the extrinsic variance of each x_i' as

$$v_i' = \mathrm{MMSE}(x_i'|\boldsymbol{b}'_{\sim i}) = \mathrm{E}\left[\left|x_i' - \mathrm{E}[x_i'|\boldsymbol{b}'_{\sim i}]\right|^2\right], \quad (18a)$$

where the expectation is taken over the joint probability space of $\boldsymbol{x}'$ and $\boldsymbol{b}'$. We denote by v the average of v_i' over the index i. Clearly, the average variance v is a function of ρ, denoted as

$$v = \psi(\rho). \quad (18b)$$

Note that (18b) is referred to as the SINR-variance transfer function of the decoder.

Remark 4: For practical FEC codes, the output variances $\{v_i'\}$ generally vary with the index i. Then, the question is whether the average variance v is a good performance measure of the decoder output, or equivalently, whether the behavior of the estimator solely depends on v. To answer this question, we recall the estimator's signal model in (13b): $\boldsymbol{y} = \boldsymbol{DFx} + \boldsymbol{\eta}$. Let v_i be the *a priori* variance of x_i. Note that $\{v_i\}$ are obtained by interleaving $\{v_i'\}$. From Assumption 1, the messages from the decoder are uncorrelated. Thus, the *a priori* covariance matrix of $\boldsymbol{x}$ is given by diag $\{v_1, v_2, \ldots, v_J\}$. Then, the *a priori* covariance of the vector $\boldsymbol{Fx}$ is given by $\boldsymbol{F}\mathrm{diag}\{v_1, v_2, \ldots, v_J\}\ \boldsymbol{F}^{\mathrm{H}}$. Clearly, $\boldsymbol{F}\mathrm{diag}\{v_1, v_2, \ldots, v_J\}\ \boldsymbol{F}^{\mathrm{H}}$ is a circulant matrix, and its diagonal is a constant given by the average variance v. Further, due to random interleaving, $\{v_1, v_2, \ldots, v_J\}$ is a random sequence. Then, it can be shown that the off-diagonals of $\boldsymbol{F}\mathrm{diag}\{v_1, v_2, \ldots, v_J\}\ \boldsymbol{F}^{\mathrm{H}}$ tend to zero as J tends to infinity, or equivalently, the *a priori* covariance of $\boldsymbol{Fx}$ is approximately a scaled identity matrix given by $v\boldsymbol{I}$. This implies that, provided that J is sufficiently large, the performance of the LMMSE detector only depends on the average output variance v of the decoder. This justifies the use of v to characterize the output of the decoder.

We next establish a relation between the rate of the FEC code (per symbol of $\boldsymbol{x}'$) and the transfer function $\psi(\rho)$. Our result is based on the SCM [19], [20]. The detail can be found in the lemma below, with the proof given in Appendix B.

Lemma 2: Assume $\psi(\rho)$ satisfies the following regularity conditions:

(i) $\psi(0) = 1$ and $\psi(\rho) \geq 0$, for $\rho \in [0, \infty)$; (19a)

(ii) monotonically decreasing in $\rho \in [0, \infty)$; (19b)

(iii) continuous and differentiable everywhere in $[0, \infty)$ except for a countable set of values of ρ; (19c)

(iv) $\lim_{\rho\to\infty} \rho\psi(\rho) = 0$. (19d)

Let Γ_n be an n-layer SCM code with the SINR-variance transfer function $\psi_n(\rho)$ and the rate R_n. Then, there exist $\{\Gamma_n\}$ such that: (i) $\psi_n(\rho) \leq \psi(\rho)$, for any $\rho \geq 0$ and any integer n; (ii) as $n \to \infty$,

$$R_n \to R = \int_0^\infty \left(\rho + \psi(\rho)^{-1}\right)^{-1} \mathrm{d}\rho. \quad (20)$$

We now give an intuitive explanation of Lemma 2. Let the MMSE of each x_i' after decoding be

$$mmse(\rho) = \mathrm{MMSE}(x_i'|\boldsymbol{b}') = \mathrm{E}\left[\left|x_i' - \mathrm{E}[x_i'|\boldsymbol{b}']\right|^2\right], \quad (21a)$$

where the expectation is taken over the joint probability space of $\boldsymbol{x}'$ and $\boldsymbol{b}'$. Compared with (18), the MMSE in (21a) is obtained by including the contribution of b_i' in estimating x_i'. We next establish a connection between $\psi(\rho)$ and $mmse(\rho)$ based on the following two hypotheses: (i) x_i' is Gaussian, and (ii) the extrinsic message of x_i' is also Gaussian. With these hypotheses, we conclude that the conditional distribution of x_i' given $\boldsymbol{b}'_{\sim i}$ is also Gaussian and given by

$$p(x_i'|\boldsymbol{b}'_{\sim i}) = \mathcal{CN}(\mathrm{E}[x_i'|\boldsymbol{b}'_{\sim\, i}], \psi(\rho)). \quad (21b)$$

Then, $mmse(\rho)$ and $\psi(\rho)$ are related as

$$\begin{aligned} mmse(\rho) &= \mathrm{MMSE}(x_i'|\boldsymbol{b}') \\ &= \mathrm{MMSE}(x_i'|b_i', \boldsymbol{b}'_{\sim i}) \\ &= \mathrm{MMSE}(x_i'|b_i', x_i' \sim p(x_i'|\boldsymbol{b}'_{\sim i})) \\ &= \mathrm{MMSE}(x_i'|b_i', x_i' \sim \mathcal{CN}(\mathrm{E}[x_i'|\boldsymbol{b}'_{\sim i}], \psi(\rho))) \\ &= \left(\rho + \psi(\rho)^{-1}\right)^{-1}. \end{aligned}$$

where the third equality follows from the fact that $b_i' \to x_i' \to \boldsymbol{x}'_{\sim i} \to \boldsymbol{b}'_{\sim i}$ forms a Markov chain (from Assumption 2 in Section II.C), the fourth equality from (21b), and the last one utilizes the fact that b_i' can be modeled as an AWGN observation in Lemma 1. From [14] (see [14, Lemma 1]), we express the code rate as

$$R = \int_0^\infty mmse(\rho)\mathrm{d}\rho = \int_0^\infty \left(\rho + \psi(\rho)^{-1}\right)^{-1} \mathrm{d}\rho \quad (22)$$

which is exactly the rate limit given in Lemma 2. Recall that the above derivation is based on two Gaussian hypotheses. These hypotheses are difficult to meet exactly, as practical systems mostly employ discrete signaling. However, the coincidence of (20) and (22) implies that SCM can approximate Gaussian signaling as the number of SCM layers tends to infinity, and therefore, the rate in (22) is indeed approachable by properly constructing an SCM-based FEC code.

E. SINR-Variance Transfer Chart

From the previous discussions, the LMMSE estimator can be characterized by $\rho = \phi(v)$; and similarly, the decoder can be characterized by $v = \psi(\rho)$. The iterative process of the estimator and decoder can be tracked by the recursion of ρ and v. Let q be the iteration number. We have

$$\rho^{(q)} = \phi(v^{(q-1)}) \text{ and } v^{(q)} = \psi(\rho^{(q)}), \quad q = 1, 2, \ldots$$

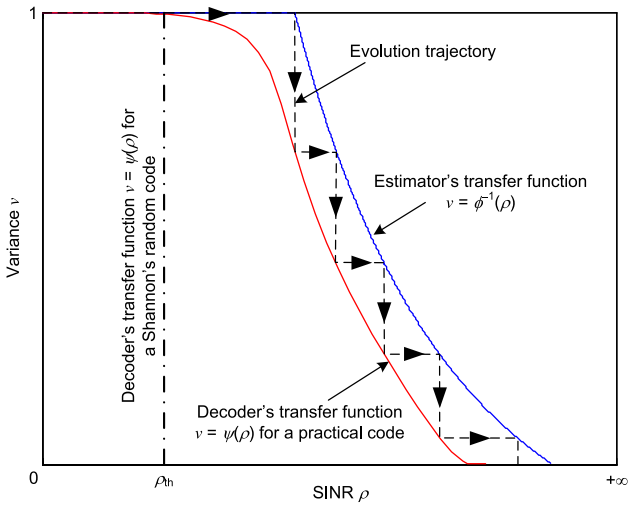

Fig. 2. An illustration of the SINR-variance transfer chart.

The recursion continues and converges to a point v^* satisfying

$$\phi(v^*) = \psi^{-1}(v^*) \text{ and } \phi(v) > \psi^{-1}(v), \quad \text{for } v \in (v^*, 1]$$

where $\psi^{-1}(\cdot)$ is the inverse of $\psi(\cdot)$ which exists since $\psi(\cdot)$ is continuous and monotonic [43]. Note that $v \leq 1$ since the signal power is normalized; $v^* = 0$ implies that $\boldsymbol{x}$ can be perfectly recovered. The above recursive process is illustrated by the SINR-variance transfer chart in Fig. 2.

We say that the estimator and the decoder are matched if

$$\phi(v) = \psi^{-1}(v), \text{ for } v \in (0, 1]. \tag{23}$$

Note that: $\phi(1) > 0$ since the estimator's output always contains the information from the channel even if there is no information from the decoder; and $\phi(0) < \infty$ since the estimator's extrinsic output cannot resolve the uncertainty introduced by the channel noise even if the messages from the decoder are perfectly reliable. Then, we equivalently express the curve-matching condition (23) as

$$\psi(\rho) = \phi^{-1}(\phi(1)) = 1, \text{ for } 0 \leq \rho < \phi(1); \tag{24a}$$
$$\psi(\rho) = \phi^{-1}(\rho), \text{ for } \phi(1) \leq \rho < \phi(0); \tag{24b}$$
$$\psi(\rho) = 0, \text{ for } \phi(0) \leq \rho < \infty. \tag{24c}$$

The above curve-matching principle plays an important role in establishing the area theorems, as seen in the next section.

IV. Area Theorems and Precoder Optimization

In this section, we establish area theorems for the proposed LP-LMMSE scheme. We further discuss the optimization of the power matrix $\boldsymbol{W}$.

A. Area Property for Unconstrained Signaling

The main theorem of this paper is presented below, with the proof given in Appendix C.

Theorem 1: As $J \to \infty$, if the detector and the decoder are matched, then an achievable rate (per antenna per channel use) of the LP-LMMSE scheme is given by

$$R = \frac{1}{N} \sum_{n=1}^{N} \log\left(1 + \frac{d_n^2}{\sigma^2}\right). \tag{25}$$

Remark 5: The curve-matching condition (23) is assumed in achieving the rate in (25); see the proof in Appendix C. Then, a fundamental question is whether this curve-matching condition can be fulfilled for any channel realization. The answer is affirmative. To see this, we recall from Lemma 2 that an SCM code can be constructed to match an arbitrary function $\psi(\rho)$ satisfying the regularity conditions in (19) and to achieve the rate in (20). Then, with (23), the question can be rephrased as: for any channel realization, does the matched $\psi(\rho)$ in (24) satisfy the regularity conditions in (19)? With $\phi(v)$ given by (15) and the properties of $\phi(v)$ in Remark 2, it is straightforward to verify that the regularity conditions in (19) are indeed satisfied for any channel realization. Note that a typical realization of $\phi^{-1}(\rho)$ is illustrated in Fig. 2.

Remark 6: We remark that (25) is the input-output mutual information of the channel in (13b) with Gaussian inputs. This observation implies that our proposed scheme is potentially information lossless. In addition, the rate R in (25) is achieved using the curve-matching SCM code with the number of SCM layers tends to infinity. Later, we will show by numerical results that, for the proposed LP-LMMSE scheme, practical values of R can be closely approached using a small number of SCM layers.

B. Water-Filling Precoding

Now we consider the optimization of the power matrix $\boldsymbol{W}$. Recall from (13a) that $\boldsymbol{D} = \boldsymbol{\Lambda}\boldsymbol{W}^{1/2}$, or equivalently, $d_i = \lambda_i w_i^{1/2}$, for $i = 1, \ldots, N$. Then, (25) becomes

$$R = \frac{1}{N} \sum_{n=1}^{N} \log\left(1 + \frac{\lambda_n^2 w_n}{\sigma^2}\right). \tag{26}$$

We aim at maximizing the above rate over $\boldsymbol{W}$ subject to the power constraint in (11). Clearly, the solution coincides with the renowned water-filling power allocation [44].

Corollary 1: Under the conditions of Theorem 1, if $\boldsymbol{W}$ follows water-filling power allocation, then the LP-LMMSE scheme achieves the water-filling capacity of the channel in (8a).

Corollary 1 shows that linear precoding, together with proper error-control coding and iterative LMMSE detection, can potentially achieve the water-filling channel capacity. This achievability is based on a single code with a matched decoding transfer function.

C. Area Property for Constrained Signaling

So far, we have shown that the proposed linear precoding scheme is asymptotically optimal (in the sense of achieving the water-filling capacity) as the signaling approaches Gaussian. However, in practice, the signaling is usually constrained on a finite-size discrete constellation. We next establish an area theorem for constrained signaling.

Define the γ-function as

$$\gamma(\rho) = \mathrm{E}\left[|x - \mathrm{E}[x|x+\eta]|^2\right] \tag{27}$$

where x is uniformly taken over $\mathcal{S}$, and η is independently drawn from $\mathcal{CN}(0, 1/\rho)$. We have the following result with the proof given in Appendix D.

Theorem 2: Suppose that the detector's inputs $\{a_i\}$ are modeled as independent observations of $\{x_i\}$ from an effective AWGN channel. Then, as $J \to \infty$, if the detector and the decoder are matched, an achievable rate of the LP-LMMSE scheme is

$$R = \log|\mathcal{S}| - \int_0^{+\infty} \gamma(\rho + \phi(\gamma(\rho)))\mathrm{d}\rho \tag{28}$$

where the ϕ-function is given by (15), and γ is defined in (27).

Remark 7: Theorem 2 holds under the AWGN assumption on $\{a_i\}$. It was observed that this assumption is empirically true for BPSK, QPSK, and SCM (with BPSK/QPSK layers). However, this assumption may be far from true if other modulation techniques, such as bit-interleaved coded modulation [47], are employed. In this later case, the area theorem based on the measure of mean-square error (MSE) established in [48] can be used for performance evaluation. We refer interested readers to [48] for details.

Remark 8: Theorem 2 can be readily extended to the case of discrete non-uniform input of system (1). The only difference is that, in the definition of the γ-function in (27), the distribution of x is replaced by a non-uniform distribution over $\mathcal{S}$. Correspondingly, "$\log|\mathcal{S}|$" in (28) is replaced by the entropy of x calculated based on the non-uniform distribution.

Remark 9: Theorem 1 and Corollary 1 reveal that our proposed linear precoding and iterative LMMSE detection scheme is capacity-achieving. It is noteworthy that there are other approaches to achieve the channel capacity. One simple approach is to use SVD to convert the MIMO channel into a set of independent parallel channels, and then to conduct water-filling power allocation over the parallel channels. It is desirable to compare the performance and complexity of our proposed scheme with the SVD-MWF scheme in practical systems with non-Gaussian inputs. Theorem 2 can be used in this comparison. Later, we will numerically demonstrate that, in this case, our proposed scheme considerably outperforms the SVD-MWF scheme.

D. Precoder Optimization for Constrained Signaling

Recall that, for Gaussian signaling, the optimal power matrix $\boldsymbol{W}$ is the water-filling solution. However, water-filling is not necessarily optimal for discrete signaling. Based on the area property in Theorem 2, we can optimize $\boldsymbol{W}$ to maximize the achievable rate of the LP-LMMSE scheme. This problem can be formulated as:

$$\text{maximize} \quad \log|\mathcal{S}| - \int_0^{+\infty} \gamma(\rho + \phi(\gamma(\rho)))\mathrm{d}\rho \tag{29a}$$

$$\text{subject to} \quad N^{-1}\sum\nolimits_{n=1}^{N} w_n \le P. \tag{29b}$$

In the above, P is the maximum signal power given in (8c). The following result is useful in solving (29).

Lemma 3: The rate R in (28) is a concave function of $\{w_1, \ldots, w_N\}$, provided that $\gamma(\rho)$ is a convex function of ρ.

The proof of Lemma 3 can be found in Appendix E. The γ-function is convex for most commonly used signaling constellations [43]. From Lemma 3, the problem in (29) can be solved using standard convex programming [49].

V. Extensions to MIMO Channels With CSIT Uncertainty

In this section, we extend the results in Sections III and IV to the general case that channel state information is not perfectly known at transmitter.

A. Linear Precoding With Imperfect CSIT

We now consider MIMO systems with partial CSIT. Here, partial CSIT means that the channel matrices are not exactly known, instead, only the statistics of the channel is known at the transmitter side. The precoding technique in Section III.B requires perfect CSIT, and so cannot be directly applied here. The following is a modified solution.

Return to the extended system in (8a)

$$\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{H}}\tilde{\boldsymbol{x}} + \tilde{\boldsymbol{\eta}}, \tag{30a}$$

where the extended channel contains J/N channel realizations:

$$\tilde{\boldsymbol{H}} = \begin{bmatrix} \boldsymbol{H}_1 & & & \\ & \boldsymbol{H}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{H}_{J/N} \end{bmatrix}. \tag{30b}$$

We focus on the ergodic case, in which $\{\boldsymbol{H}_i\}$ are independent realizations of an M-by-N random matrix, with abuse of notation, denoted by $\boldsymbol{H}$. This model also includes orthogonal frequency-division multiplexing (OFDM) systems [40] with independent fading over different sub-carriers.

The channel input $\tilde{\boldsymbol{x}}$ is related to $\boldsymbol{x}$ as

$$\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{P}}\tilde{\boldsymbol{\Pi}}\tilde{\boldsymbol{F}}\boldsymbol{x} \tag{31a}$$

with

$$\tilde{\boldsymbol{P}} = \mathrm{diag}\{\boldsymbol{P}, \boldsymbol{P}, \ldots, \boldsymbol{P}\} \tag{31b}$$

$$\tilde{\boldsymbol{F}} = \mathrm{diag}\{\boldsymbol{F}, \boldsymbol{F}, \ldots, \boldsymbol{F}\}. \tag{31c}$$

where $\tilde{\boldsymbol{P}}$ is a J-by-J block-diagonal matrix with $\boldsymbol{P}$ of size N-by-N, $\tilde{\boldsymbol{\Pi}}$ is a J-by-J permutation matrix, and $\tilde{\boldsymbol{F}}$ is a J-by-J block- diagonal matrix with each block $\boldsymbol{F}$ being the normalized DFT matrix of size L-by-L. Combining (30a) and (31) and letting $\boldsymbol{y} = \tilde{\boldsymbol{y}}$ and $\boldsymbol{\eta} = \tilde{\boldsymbol{\eta}}$, we obtain an equivalent channel as

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\eta} \tag{32a}$$

with

$$\boldsymbol{A} = \tilde{\boldsymbol{H}}\tilde{\boldsymbol{P}}\tilde{\boldsymbol{\Pi}}\tilde{\boldsymbol{F}}. \tag{32b}$$

The power constraint now becomes

$$N^{-1}\mathrm{tr}\{\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}\} \le P. \tag{33}$$

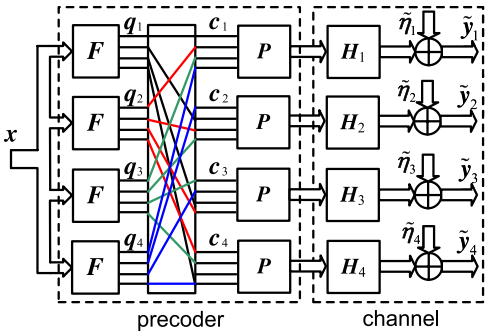

Fig. 3. An illustration of the proposed precoder for imperfect CSIT with $J = 16$ and $N = 4$.

The precoding matrix $\tilde{\boldsymbol{P}} = \text{diag}\{\boldsymbol{P}, \ldots, \boldsymbol{P}\}$ allows the precoder to exploit the benefit provided by the available CSIT.[4] The optimization of $\boldsymbol{P}$ is briefly discussed in Section V.C.

The precoding matrix $\tilde{\boldsymbol{F}}$ ensures that every coded symbol in $\boldsymbol{x}$ is sufficiently dispersive over time and space. Thus, it is required that the DFT size L is sufficiently large. A convenient choice of L is $L = J/N$, and then there are N length-L DFT matrices in $\tilde{\boldsymbol{F}}$.

Now we consider the design of the permutation matrix $\tilde{\boldsymbol{\Pi}}$. Let $\boldsymbol{q} = [\boldsymbol{q}_1^{\mathrm{T}}, \ldots, \boldsymbol{q}_N^{\mathrm{T}}]^{\mathrm{T}} = \tilde{\boldsymbol{F}}\boldsymbol{x}$ and $\boldsymbol{x} = [\boldsymbol{x}_1^{\mathrm{T}}, \ldots, \boldsymbol{x}_N^{\mathrm{T}}]^{\mathrm{T}}$, where each DFT block $\boldsymbol{q}_i = \boldsymbol{F}\boldsymbol{x}_i$ is of size L-by-1. Also denote $\boldsymbol{c} = [\boldsymbol{c}_1^{\mathrm{T}}, \ldots, \boldsymbol{c}_{J/N}^{\mathrm{T}}]^{\mathrm{T}} = \tilde{\boldsymbol{\Pi}}\boldsymbol{q}$, where each $\boldsymbol{c}_i$ is an N-by-1 vector. The two criteria for $\tilde{\boldsymbol{\Pi}}$ are listed below.

(i) For each i, the entries of $\boldsymbol{q}_i$ are transmitted at different channel uses, i.e., no two entries of $\boldsymbol{q}_i$ are connected to a same $\boldsymbol{c}_j$ for any index $j = 1, \ldots, J/N$; and

(ii) Treat $\boldsymbol{P}\boldsymbol{H}_k$ as the kth realization of the equivalent channel $\boldsymbol{P}\boldsymbol{H}$. Then, any N consecutive entries of $\boldsymbol{q}_i$ are transmitted at different *transmit-antennas* of the equivalent channel $\boldsymbol{P}\boldsymbol{H}$. As a result, for each index j, the set of {the jth entry of $\boldsymbol{c}_k | k = 1, \ldots, J/N$} contains J/N^2 entries of $\boldsymbol{q}_i, i = 1, \ldots, N$.[5]

The choice of $\tilde{\boldsymbol{\Pi}}$ satisfying the above criteria is not unique. A simple choice is as follows: for each index k, the kth entry of $\boldsymbol{q}_1$ is connected to the (k mod N)th input of the equivalent channel $\boldsymbol{P}\boldsymbol{H}_k$; then, for $j = 2, \ldots, J/N$, the connection pattern of each $\boldsymbol{q}_j$ is just a one-entry cyclic shift of the previous one. An example for $J = 16$ and $N = 4$ is illustrated in Fig. 3. The above choice of $\tilde{\boldsymbol{F}}$ and $\tilde{\boldsymbol{\Pi}}$ ensures that the behavior of the iterative receiver can be characterized by a single-variable recursion, as will be detailed in the next subsection.

B. SINR-Variance Transfer Chart

Iterative LMMSE detection described in Section II is applied to the system in (32). We next show that the outputs of the detector can still be characterized by a single SINR value.

[4]We emphasize that $\boldsymbol{P}$ is designed to be adaptive to the available channel state information. Here, $\boldsymbol{P}$ remains constant for different channel uses, as the channel statistics does not change. However, if the channel statistics varies, $\boldsymbol{P}$ should vary accordingly.

[5]Here we assume that J is properly chosen such that J/N^2 is an integer.

With abuse of notation, define

$$\phi(v) = \left(\frac{1}{N}\mathrm{E}\left[\mathrm{tr}\left\{\left(\frac{1}{v}\boldsymbol{I} + \frac{1}{\sigma^2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^{\mathrm{H}}\right)^{-1}\right\}\right]\right)^{-1} - \frac{1}{v}. \quad (34)$$

where $\boldsymbol{Q} = \boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}$, and the expectation is taken over the distribution of $\boldsymbol{H}$. We aim to show that

$$\rho_i = 1/u_i = \phi(v), \quad \text{for } i = 1, \ldots, J. \quad (35)$$

With (32b), we can rewrite (5c) as

$$\begin{aligned}\boldsymbol{M} &= \tilde{\boldsymbol{F}}^{\mathrm{H}}\tilde{\boldsymbol{\Pi}}^{\mathrm{H}}\big(v\boldsymbol{I} - v^2\tilde{\boldsymbol{P}}^{\mathrm{H}}\tilde{\boldsymbol{H}}^{\mathrm{H}}(v\tilde{\boldsymbol{H}}\tilde{\boldsymbol{P}}\tilde{\boldsymbol{P}}^{\mathrm{H}}\tilde{\boldsymbol{H}}^{\mathrm{H}} + \sigma^2\boldsymbol{I})^{-1}\tilde{\boldsymbol{H}}\tilde{\boldsymbol{P}}\big)\tilde{\boldsymbol{\Pi}}\tilde{\boldsymbol{F}} \\ &= \tilde{\boldsymbol{F}}^{\mathrm{H}}\boldsymbol{B}\tilde{\boldsymbol{F}}\end{aligned}$$

where

$$\boldsymbol{B} = \tilde{\boldsymbol{\Pi}}^{\mathrm{H}}\big(v\boldsymbol{I} - v^2\tilde{\boldsymbol{P}}^{\mathrm{H}}\tilde{\boldsymbol{H}}^{\mathrm{H}}(v\tilde{\boldsymbol{H}}\tilde{\boldsymbol{P}}\tilde{\boldsymbol{P}}^{\mathrm{H}}\tilde{\boldsymbol{H}}^{\mathrm{H}} + \sigma^2\boldsymbol{I})^{-1}\tilde{\boldsymbol{H}}\tilde{\boldsymbol{P}}\big)\tilde{\boldsymbol{\Pi}}. \quad (36)$$

We can express $\boldsymbol{M}$ and $\boldsymbol{B}$ in a block-wise form with each block of size (J/N)-by-(J/N). Let $\boldsymbol{M}_i$ and $\boldsymbol{B}_i$ be the ith diagonal block of $\boldsymbol{M}$ and $\boldsymbol{B}$, respectively. Then, we obtain

$$(\boldsymbol{M})_{\text{diag}} = \begin{bmatrix}(\boldsymbol{F}^{\mathrm{H}}\boldsymbol{B}_1\boldsymbol{F})_{\text{diag}} & & \\ & \ddots & \\ & & (\boldsymbol{F}^{\mathrm{H}}\boldsymbol{B}_N\boldsymbol{F})_{\text{diag}}\end{bmatrix},$$

where $(\cdot)_{\text{diag}}$ returns a diagonal matrix specified by the diagonal of the matrix in the parenthesis. Recall that both $\tilde{\boldsymbol{H}}$ and $\tilde{\boldsymbol{P}}$ are block diagonal, and so is $\tilde{\boldsymbol{P}}^{\mathrm{H}}\tilde{\boldsymbol{H}}^{\mathrm{H}}(v\tilde{\boldsymbol{H}}\tilde{\boldsymbol{P}}\tilde{\boldsymbol{P}}^{\mathrm{H}}\tilde{\boldsymbol{H}}^{\mathrm{H}} + \sigma^2\boldsymbol{I})^{-1}\tilde{\boldsymbol{H}}\tilde{\boldsymbol{P}}$. Thus, from the criterion (i) of $\tilde{\boldsymbol{\Pi}}$ in Section V.A, every $\boldsymbol{B}_i$ is diagonal. Therefore, as $J \to \infty$

$$\begin{aligned}M_i(k,k) &= \frac{N}{J}\mathrm{tr}\{\boldsymbol{B}_i\} && (37a)\\ &\to \frac{1}{J}\mathrm{tr}\{\boldsymbol{B}\} && (37b)\\ &= \frac{1}{J}\mathrm{tr}\left\{v\boldsymbol{I} - v^2\tilde{\boldsymbol{P}}^{\mathrm{H}}\tilde{\boldsymbol{H}}^{\mathrm{H}}(v\tilde{\boldsymbol{H}}\tilde{\boldsymbol{P}}\tilde{\boldsymbol{P}}^{\mathrm{H}}\tilde{\boldsymbol{H}}^{\mathrm{H}} + \sigma^2\boldsymbol{I})^{-1}\tilde{\boldsymbol{H}}\tilde{\boldsymbol{P}}\right\} && (37c)\\ &= \frac{1}{J}\mathrm{tr}\left\{\left(\frac{1}{v}\boldsymbol{I} + \frac{1}{\sigma^2}\tilde{\boldsymbol{H}}\tilde{\boldsymbol{P}}\tilde{\boldsymbol{P}}^{\mathrm{H}}\tilde{\boldsymbol{H}}^{\mathrm{H}}\right)^{-1}\right\} && (37d)\\ &= \frac{1}{J}\sum_{i=1}^{J/N}\mathrm{tr}\left\{\left(\frac{1}{v}\boldsymbol{I} + \frac{1}{\sigma^2}\boldsymbol{H}_i\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{H}_i^{\mathrm{H}}\right)^{-1}\right\} && (37e)\\ &\to \frac{1}{N}\mathrm{E}\left[\mathrm{tr}\left\{\left(\frac{1}{v}\boldsymbol{I} + \frac{1}{\sigma^2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^{\mathrm{H}}\right)^{-1}\right\}\right] && (37f)\end{aligned}$$

where (37a) follows from the fact that, for each i

$$(\boldsymbol{F}^{\mathrm{H}}\boldsymbol{B}_i\boldsymbol{F})_{\text{diag}} = \left(\frac{N}{J}\sum_{n=1}^{N}B_i(n,n)\right)\boldsymbol{I} = \frac{N}{J}\mathrm{tr}\{\boldsymbol{B}_i\}\boldsymbol{I},$$

(37b) from the criterion (ii) of $\tilde{\boldsymbol{\Pi}}$, (37c) from (36), and (37e) by substituting $\tilde{\boldsymbol{H}} = \text{diag}\{\boldsymbol{H}_1, \ldots, \boldsymbol{H}_{J/N}\}$, $\tilde{\boldsymbol{P}} = \text{diag}\{\boldsymbol{P}, \ldots, \boldsymbol{P}\}$, and $\boldsymbol{Q} = \boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}$. Substituting (37) into (5a), we arrive at (35).

Now, similar to Lemma 1, we have the following result.

Lemma 4: The detector's output b_i can be expressed as $b_i = x_i + n_i$, where n_i is independent of x_i and converges to $\mathcal{CN}(0, 1/\phi(v))$ as $J \to \infty$.

The above lemma is literally the same as Lemma 1, except that $\phi(v)$ here is given by (34). The proof mostly follows that of Lemma 1. We omit the details for brevity. With Lemma 4, we can still characterize the behavior of the iterative receiver using the SINR-variance transfer functions, namely, $\rho = \phi(v)$ and $v = \psi(\rho)$, as described in Section III.E.

C. Area Theorem and Precoder Optimization

Now we are ready to present the following result.

Theorem 3: As $J \to \infty$, if the detector and the decoder are matched, an achievable information rate (per transmit antenna per channel use) of the LP-LMMSE scheme is given by

$$R = \frac{1}{N}\mathrm{E}\left[\log\det\left(\boldsymbol{I} + \frac{1}{\sigma^2}\boldsymbol{HQH}^{\mathrm{H}}\right)\right]. \tag{38}$$

The proof of Theorem 3 is given in Appendix F.

With Theorem 3, we can formulate the following rate-maximization problem:

$$\text{maximize } \mathrm{E}\left[\log\det\left(\boldsymbol{I} + \frac{1}{\sigma^2}\boldsymbol{HQH}^{\mathrm{H}}\right)\right] \tag{39a}$$

$$\text{subject to } N^{-1}\mathrm{tr}\{\boldsymbol{Q}\} \le P. \tag{39b}$$

The above optimization problem is convex, and thus can be solved numerically using standard convex programming tools. Moreover, the explicit solutions to (39) in a variety of CSIT settings have been studied in the literature. We refer interested readers to [34], [50], and [51] for more details.

For constrained signaling, it is straightforward to show that Theorem 2 holds literally for the partial CSIT case, except that $\phi(v)$ is replaced by (34). Then, we can formulate an optimization problem similar to (29), which is solvable using standard convex programming. Details are omitted for simplicity.

VI. Numerical Results

In this section, we provide numerical examples to demonstrate the achievable rates of the proposed scheme. Note that the channel SNR is defined as $\mathrm{SNR} = PN/\sigma^2$.

We first consider the case of full CSIT in a randomly generated 2×2 MIMO 3-tap ISI channel with the tap coefficients

$$\begin{bmatrix} 0.5339 + \mathrm{j}0.5395 & -0.4245 + \mathrm{j}0.0648 \\ -0.3347 - \mathrm{j}0.3727 & -0.4672 - \mathrm{j}0.2420 \end{bmatrix}, \tag{40a}$$

$$\begin{bmatrix} 0.0582 - \mathrm{j}0.2706 & 0.1525 - \mathrm{j}0.7565 \\ -0.4968 - \mathrm{j}0.1543 & -0.5243 - \mathrm{j}0.5915 \end{bmatrix}, \tag{40b}$$

and

$$\begin{bmatrix} -0.5262 - \mathrm{j}0.2654 & -0.3714 - \mathrm{j}0.2865 \\ 0.6721 - \mathrm{j}0.1635 & 0.1607 - \mathrm{j}0.2695 \end{bmatrix}. \tag{40c}$$

The DFT is applied to convert the above MIMO ISI channel to a set of $J = 256$ 2-by-2 parallel MIMO channels. Note that the effect of cyclic prefix is ignored here.

Fig. 4 shows the achievable rates of the proposed LP-LMMSE scheme with various power allocation strategies. Various power allocation policies are considered, including equal power (EP) allocation, water-filling (WF), and optimized power (OP) allocation. For EP, $\boldsymbol{W} = \varepsilon\boldsymbol{I}$ where ε is a scaling factor to meet the power budget; for WF, $\boldsymbol{W}$ is determined by the water-filling solution; and for OP, $\boldsymbol{W}$ is obtained by solving (29). QPSK modulation is employed.

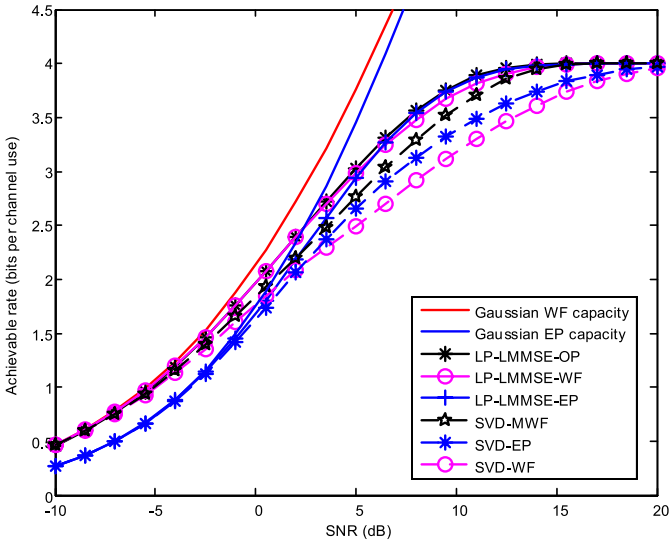

Fig. 4. The achievable rates of the proposed LP-LMMSE scheme and the SVD-based parallel transmission scheme with QPSK modulation in the MIMO-ISI channel in (40). The water-filling capacity and Gaussian EP capacity are also included for comparison.

From Fig. 4, we see that the LP-LMMSE schemes with OP and WF power allocation have similar performance, and both considerably outperform the flat precoder in the low SNR region. This implies that the water-filling precoder provides an attractive low-complexity near-optimum option for discrete signaling. Fig. 4 also includes the Gaussian water-filling capacity and Gaussian EP (i.e., i.i.d.) capacity as references. We see that the LP-LMMSE schemes with OP and WP approach the water-filling capacity in the low SNR region. Further, we compare the performance of the LP-LMMSE scheme with that of the parallel transmission scheme (in which the MIMO channels are converted into parallel independent scalar channels using SVD) with various power allocation strategies. We see that the LP-LMMSE scheme outperforms the parallel transmission scheme with MWF considerably in the medium SNR range. This indicates the advantage of linear precoding over parallel independent transmission for MIMO channels.

In Fig. 5, we employ both QPSK modulation and standard 16-QAM (see [57, SCM-1]) for signaling. Fig. 5 demonstrates that the achievable rate of the LP-LMMSE scheme is significantly increased by changing the signaling constellation from QPSK to 16-QAM. We see that the gap between the achievable rate of LP-LMMSE with 16-QAM and the water-filling capacity is not significant for a rate up to 4 bits. For a higher transmission rate, a larger signaling constellation is necessary.

We now provide code design examples to verify that the theoretical limits given by our analysis are indeed approachable. We basically follow the design approach described in [45]. Here, we only provide the design results. We consider standard 16-QAM signaling with 2-layer SCM. Each SCM

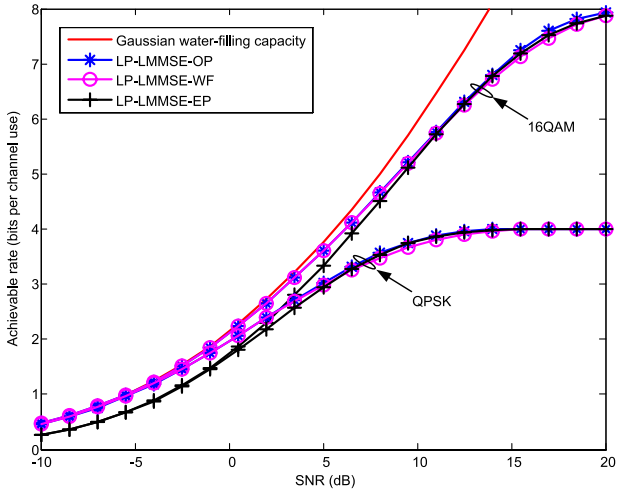

Fig. 5. The achievable rates of the proposed LP-LMMSE scheme in the MIMO-ISI channel in (40). Both QPSK and standard 16-QAM are considered. The water-filling capacity is also included for reference.

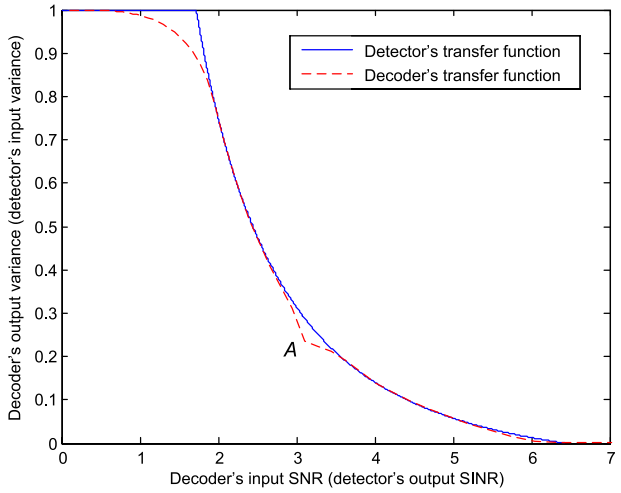

Fig. 6. The SINR-variance transfer functions of the detector and the decoder for the LP-LMMSE scheme with 2-layer SCM over the MIMO ISI channel in (40). Water-filling precoding and standard 16-QAM constellation are used. The transfer function of the detector is given by (15) at the channel SNR = 6.7 dB.

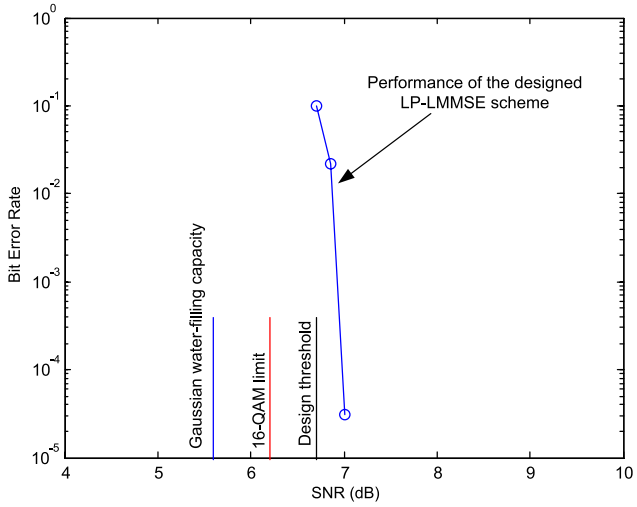

Fig. 7. The BER performance of the LP-LMMSE scheme with 2-layer SCM over the MIMO ISI channel in (40). Water-filling precoding and standard 16-QAM constellation are used. Some simulation settings are $M = N = 2$, $J = 256$, and $K = 256$.

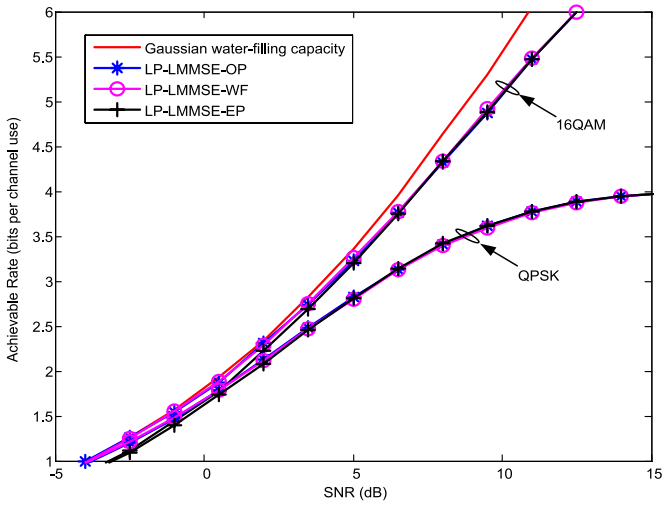

Fig. 8. The achievable rates of the LP-LMMSE scheme in the flat-fading 2-by-2 MIMO channel with partial CSIT (of $\theta = 0.5$). The water-filling capacity is also included for reference.

layer is QPSK modulated with Gray mapping. (Note: for standard 16-QAM signaling, the power ratio of the two SCM layers is 4:1.) Two irregular LDPC codes are designed based on the curve-matching principle. For the SCM layer with higher power, the edge degree distributions of the LDPC code are given by $\{\lambda_1(x) = 0.3495x + 0.2142x^2 + 0.1165x^6 + 0.0857x^7 + 0.1331x^{17} + 0.1010x^{20}, \rho_1(x) = x^6\}$; for the other layer, the edge degree distributions of the LDPC code are $\{\lambda_2(x) = 0.3907x + 0.1854x^2 + 0.0968x^{10} + 0.1877x^{11} + 0.1394x^{33}, \rho_2(x) = x^6\}$. The system throughput is 4 bits per channel use. The transfer curves of the detector and the decoder are illustrated in Fig. 6, and the simulated system performance is given in Fig. 7. From Fig. 7, the design threshold is 6.7 dB, 1.1 dB away from the water-filling capacity, 0.5 dB away from the performance limit of the LP-LMMSE scheme with standard 16-QAM. The simulated system performance with code length $\approx 10^5$ at BER $= 10^{-4}$ is about 7.0 dB, only 0.3 dB away from the design threshold.

It is interesting to pay special attention to point A in Fig. 6. This is the transition point where, for the two layers of the LDPC codes, the one with higher power is nearly fully decoded (i.e., with very reliable outputs) and the one with lower power starts to be decoded.

Now consider the case of partial CSIT. We assume the mean- feedback model [53], i.e., each $\boldsymbol{H}_i$ is a random realization of

$$\boldsymbol{H} = \mathrm{E}[\boldsymbol{H}] + \Delta\boldsymbol{H}$$

where the entries of $\mathrm{E}[\boldsymbol{H}]$ are independently drawn from $\mathcal{CN}(0, \theta)$ with $\theta \in [0, 1]$, and those of $\Delta\boldsymbol{H}$ from $\mathcal{CN}(0, 1-\theta)$. In simulation, θ is set to 0.5. The channel mean $\mathrm{E}[\boldsymbol{H}]$ remains constant for each frame, but varies independently from frame to frame. Each frame contains 50 independently generated $\Delta\boldsymbol{H}$. The achievable rates of the LP-LMMSE scheme (averaged over 2000 frames) are illustrated in Fig. 8. We see that similar trends as in Fig. 5 have been observed.

VII. Conclusion

In this paper, we proposed the LP-LMMSE scheme for coded linear systems, and established analytical techniques to evaluate the achievable rate of the proposed scheme with multi-ary signaling. The main results are summarized as follows.

We first discussed the case of full CSIT. We showed that the LP-LMMSE scheme can be analyzed using the SINR-variance transfer chart. We established an area theorem and showed that the proposed LP-LMMSE scheme is information lossless for unconstrained signaling. We also developed an area theorem to evaluate the achievable rate of the LP-LMMSE scheme with finite discrete signaling. Based on these area theorems, the precoder optimization is discussed and the optimal precoders were developed. It was shown that a properly designed LP-LMMSE scheme can achieve the water-filling channel capacity.

We then considered the case of partial CSIT. We showed that, with the modified linear precoding technique, our results for the case of full CSIT readily carry over to the case of partial CSIT. Particularly, we showed that a properly designed LP-LMMSE scheme can achieve the ergodic capacity of the fading MIMO channel.

Appendix A
Proof of Lemma 1

Substituting (4b) and $\boldsymbol{A} = \boldsymbol{DF}$ into (4a), we obtain the ith entry of $\hat{\boldsymbol{x}}$ as

$$\begin{aligned}\hat{x}_i &= \bar{x}_i + v\boldsymbol{f}_i^{\mathrm{H}}\boldsymbol{D}^{\mathrm{H}}(v\boldsymbol{D}\boldsymbol{D}^{\mathrm{H}} + \sigma^2\boldsymbol{I})^{-1}(\boldsymbol{y} - \boldsymbol{A}\bar{\boldsymbol{x}}) \\ &= \left(1 - v\boldsymbol{f}_i^{\mathrm{H}}\boldsymbol{D}^{\mathrm{H}}(v\boldsymbol{D}\boldsymbol{D}^{\mathrm{H}} + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{D}\boldsymbol{f}_i\right)\bar{x}_i \\ &\quad + v\boldsymbol{f}_i^{\mathrm{H}}\boldsymbol{D}^{\mathrm{H}}(v\boldsymbol{D}\boldsymbol{D}^{\mathrm{H}} + \sigma^2\boldsymbol{I})^{-1}(\boldsymbol{y} - \boldsymbol{A}\bar{\boldsymbol{x}}_{\setminus i}) \\ &= v^{-1}M(i,i)\bar{x}_i + v\boldsymbol{f}_i^{\mathrm{H}}\boldsymbol{D}^{\mathrm{H}}(v\ \boldsymbol{D}\boldsymbol{D}^{\mathrm{H}} + \sigma^2\boldsymbol{I})^{-1} \\ &\quad \times(\boldsymbol{y} - \boldsymbol{A}\bar{\boldsymbol{x}}_{\setminus i}) \end{aligned} \tag{A1}$$

where $\bar{\boldsymbol{x}}_{\setminus i}$ represents the vector obtained by setting the ith entry of $\bar{\boldsymbol{x}}$ to zero, and the last equality follows from (14b). Then

$$\begin{aligned} b_i &= \frac{v}{M(i,i)\rho}\boldsymbol{f}_i^{\mathrm{H}}\boldsymbol{D}^{\mathrm{H}}(v\ \boldsymbol{D}\boldsymbol{D}^{\mathrm{H}} + \sigma^2\boldsymbol{I})^{-1}(\boldsymbol{y} - \boldsymbol{A}\bar{\boldsymbol{x}}_{\setminus i}) \\ &= \frac{v}{M(i,i)\rho}\boldsymbol{f}_i^{\mathrm{H}}\boldsymbol{D}^{\mathrm{H}}(v\ \boldsymbol{D}\boldsymbol{D}^{\mathrm{H}} + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{D}\boldsymbol{f}_i x_i + n_i \\ &= x_i + n_i \end{aligned}$$

where the first equality follows by substituting (A1) into (5b) and by noting $u_i = 1/\rho$, the second equality by substituting (13b) and (16), and the last equality from (14a) and (15). Moreover, it can be verified that

$$\mathrm{E}[n_i] = 0 \text{ and } \mathrm{E}[|n_i|^2] = \phi(\upsilon)^{-1}.$$

We next show that, for a sufficiently large J, the distribution of n_i converges to a normal distribution. This basically follows the central limit theorem. The details are as follows.

Recall the expression of n_i in (16). As η is Gaussian, we only need to show that

$$e_i = \boldsymbol{f}_i^{\mathrm{H}}\boldsymbol{D}^{\mathrm{H}}(v\boldsymbol{D}\boldsymbol{D}^{\mathrm{H}} + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{D}\boldsymbol{F}(\boldsymbol{x}_{\setminus i} - \bar{\boldsymbol{x}}_{\setminus i})$$

converges to a normal distribution. Let $\boldsymbol{s}$ be the diagonal of $\boldsymbol{D}^{\mathrm{H}}(v\boldsymbol{D}\boldsymbol{D}^{\mathrm{H}} + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{D}$, and let

$$\boldsymbol{t} = J^{-1/2}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{s}. \tag{A2}$$

It is straightforward to verify that

$$t_k = J^{-1/2}\boldsymbol{f}_k^{\mathrm{H}}\boldsymbol{s} = \boldsymbol{f}_i^{\mathrm{H}}\ \boldsymbol{D}^{\mathrm{H}}(v\boldsymbol{D}\boldsymbol{D}^{\mathrm{H}} + \sigma^2\boldsymbol{I})^{-1}\ \boldsymbol{D}\boldsymbol{f}_{i+k-1}.$$

With the above, e_i can be equivalently expressed as

$$e_i = \sum_{k=2}^{J} t_k\left(x_{[i+k-1]} - \bar{x}_{[i+k-1]}\right) \tag{A3}$$

where $x_{[i+k-1]} = x_{i+k-1-J}$ if $i + k - 1 > J$.

Both $\boldsymbol{x}$ and $\bar{\boldsymbol{x}}$ are $i.i.d.$ sequences, and so is $\boldsymbol{x} - \bar{\boldsymbol{x}}$. We consider the following limit:

$$\begin{aligned}\omega_\varepsilon &= \lim_{J\to\infty}\frac{\sum_{i=2}^{J}\mathrm{E}\left|t_i\left(x_{[i+k-1]} - \bar{x}_{[i+k-1]}\right)\right|^{2+\varepsilon}}{\left(\sum_{i=2}^{J}\mathrm{E}\left|t_i\left(x_{[i+k-1]} - \bar{x}_{[i+k-1]}\right)\right|^2\right)^{\frac{2+\varepsilon}{2}}} \\ &= \frac{\mathrm{E}\,|x_1 - \bar{x}_1|^{2+\varepsilon}}{\left(\mathrm{E}\,|x_1 - \bar{x}_1|^2\right)^{\frac{2+\varepsilon}{2}}}\cdot\lim_{J\to\infty}\frac{\sum_{k=2}^{J}|t_k|^{2+\varepsilon}}{\left(\sum_{k=2}^{J}|t_k|^2\right)^{\frac{2+\varepsilon}{2}}} \end{aligned} \tag{A4}$$

where ε is a positive number, and the second equality follows from the fact that

$$\mathrm{E}\left[|x_i - \bar{x}_i|^{2+\varepsilon}\right] = \mathrm{E}\left[|x_i - \mathrm{E}[x_i|a_i]|^{2+\varepsilon}\right] = c_\varepsilon, \text{ for any } i,$$

with c_ε invariant to the index i. From the Lyapunov's central limit theorem, if $\omega_\varepsilon = 0$ for some $\varepsilon > 0$, e_i in (A3) converges in distribution to a normal distribution ([54, p. 309]).

What remains is to verify $\omega_\varepsilon = 0$ for $\varepsilon = 2$. From (A4), it suffices to show that

$$\lim_{J\to\infty}\frac{\sum_{k=2}^{J}|t_k|^4}{\left(\sum_{k=2}^{J}|t_k|^2\right)^2} = 0. \tag{A5}$$

To this end, define $\Delta\boldsymbol{s} = \boldsymbol{s} - \bar{s}\boldsymbol{1}$, where $\boldsymbol{s}$ is the diagonal of $\boldsymbol{D}^{\mathrm{H}}(v\boldsymbol{D}\boldsymbol{D}^{\mathrm{H}}+\sigma^2\boldsymbol{I})^{-1}\boldsymbol{D}$, $\bar{s}$ is the average value of the entries of $\boldsymbol{s}$, and $\boldsymbol{1}$ represents an all-one vector. From the discussions below (9), the ordering of the diagonal entries of $\boldsymbol{\Lambda}$ are chosen such that the entries of $\Delta\boldsymbol{s}$ are asymptotically uncorrelated, i.e., for any $i \neq 0$,

$$\lim_{J\to\infty}\frac{\sum_{i=1}^{J}\Delta s_i\,\Delta s_{i+i'}}{\sum_{i=1}^{J}\Delta s_i^2} = 0. \tag{A6}$$

From (A2), we can express

$$t_k = \frac{1}{J}\sum_{i=1}^{J}\Delta s_i\exp(\mathrm{j}2\pi(i-1)(k-1)/J), \text{ for } k = 2,\ldots,J.$$

Then, with some straightforward but tedious derivations, we can obtain (A5), which completes the proof of Lemma 1.

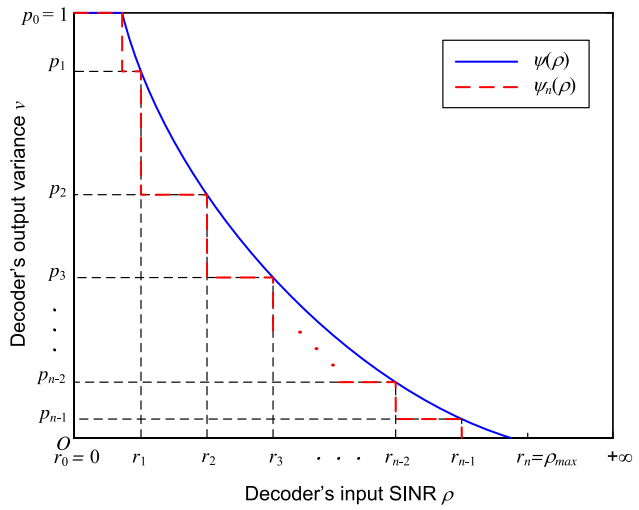

Fig. 9. An illustration of the target SINR-variance transfer curve $\psi(\rho)$ and the corresponding transfer curve $\psi_n(\rho)$ of the matched n-layer SCM code.

Appendix B
Proof of Lemma 2

We prove Lemma 2 by constructing a series of n-layer SCM code Γ_n as follows. Let $\rho_{\max}$ be an arbitrary positive number. We quantize the SNR range $(0, \rho_{\max}]$ with interval $\Delta r \equiv \rho_{\max}/n$. The quantization thresholds are given by $r_0 = 0, r_1 = \rho_{\max}/n, \ldots, r_n = \rho_{\max}$. As illustrated in Fig. 9, denote

$$p_i = \psi(r_i), \text{ for } i = 0, 1, \ldots, n. \tag{A7}$$

Note that $\Delta p_i = p_{i-1} - p_i \geq 0$ since $\psi(\rho)$ is monotonically decreasing. Let Δp_i be the signal power of the ith layer of Γ_n, for $i = 1, \ldots, n$. The ith layer of Γ_n is encoded using a random codebook drawn from $\mathcal{CN}(0, \Delta p_i)$ at a rate of

$$R_{n,i} = \log\left(1 + \frac{\Delta p_i}{p_{i+1} + r_i^{-1}}\right), \text{ for } i = 1, \ldots, n. \tag{A8}$$

Here, we assume that each SCM layer employs Gaussian signaling. The extension to discrete signaling will be given at the end of the proof. We have the following result.

Fact 1: For any $i \in \{0, 1, \ldots, n\}$ and the decoder's input SNR$\rho \in [r_i, r_{i+1})$, the first i layers of Γ_n can be successively decoded in the order from layer 1 to layer i.

Proof of Fact 1: We prove by induction. Fact 1 trivially holds for $i = 0$. Now suppose that Fact 1 holds for $i = k-1$. Consider any input SNR $\rho \in (r_k, r_{k+1}]$. By assumption, the first $k-1$ layers are decodable. After decoding and canceling the first $k-1$ layers from the decoder's input, the SINR seen by the kth layer is

$$\text{SINR} = \frac{\Delta p_k}{p_{k+1} + \rho^{-1}} > \frac{\Delta p_k}{p_{k+1} + r_k^{-1}}, \tag{A9}$$

where ρ^{-1} is the noise power (as the decoder's input SNR is ρ and the signal power is normalized to 1), and the inequality follows from $\rho \in (r_k, r_{k+1}]$. Thus, from the Shannon's capacity formula, the kth layer with rate $R_{n,i}$ in (A8) is decodable, hence the proof of Fact 1.

Now consider the SINR-variance transfer function of Γ_n, denoted by $\psi_n(\rho)$. Under APP decoding, the SINR-variance transfer function of a random code with a sufficiently large code length is given by (see [14, Th. 3])

$$\psi_{\text{random}}(\rho) = \begin{cases} p, & \text{for } \rho < \rho_{\text{th}}; \\ 0, & \text{for } \rho > \rho_{\text{th}} \end{cases}$$

where p is the transmission power and ρ_{th} is the SNR threshold stipulated by the channel capacity (cf., Fig. 2). Thus, together with Fact 1, it can be readily verified that $\psi_n(\rho)$ has a ladder-shaped form (as illustrated in Fig. 9) satisfying

$$\psi_n(\rho) \leq \psi(\rho), \text{ for any } \rho > 0.$$

What remains is to show that $R_n \to R$ as $n \to \infty$. To this end, we have

$$\begin{aligned}
\lim_{n\to\infty} R_n &= \lim_{n\to\infty} \sum_{i=1}^{n} R_{n,i} \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} \left(\frac{\Delta p_i}{p_{i+1} + r_i^{-1}} + o\left(\frac{\Delta p_i}{p_{i+1} + r_i^{-1}} \right) \right) \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} \frac{\psi(r_{i-1}) - \psi(r_i)}{\psi(r_{i+1}) + r_i^{-1}} \\
&= -\lim_{n\to\infty} \sum_{i=1}^{n} \frac{\psi'(r_i)(r_i - r_{i-1})}{\psi(r_{i+1}) + r_i^{-1}} \\
&= -\int_0^{\rho_{\max}} \frac{\rho\psi'(\rho)}{\rho\psi(\rho) + 1} \mathrm{d}\rho \\
&= \int_0^{\rho_{\max}} \frac{\psi(\rho)}{\rho\psi(\rho) + 1} \mathrm{d}\rho - \int_0^{\rho_{\max}} \frac{\rho\psi'(\rho) + \psi(\rho)}{\rho\psi(\rho) + 1} \mathrm{d}\rho \\
&= \int_0^{\rho_{\max}} \frac{\psi(\rho)}{\rho\psi(\rho) + 1} \mathrm{d}\rho - \log\left(\rho\psi(\rho) + 1\right)\Big|_{\rho=0}^{\rho=\rho_{\max}} \\
&= \int_0^{\rho_{\max}} \frac{1}{\psi(\rho)^{-1} + \rho} \mathrm{d}\rho - \log\left(\rho_{\max}\psi(\rho_{\max}) + 1\right)
\end{aligned} \tag{A10}$$

where the second equality follows from (A8), the third equality utilizes (A7), and the fourth equality utilizes the fact that $\psi(\rho)$ is first-order differentiable.

Recall that $\rho_{\max}$ is arbitrary. Letting $\rho_{\max} \to \infty$, together with (19d), we obtain from (A10) that

$$\lim_{n\to\infty} R_n = \int_0^{\infty} \left(\psi(\rho)^{-1} + \rho \right) \mathrm{d}\rho.$$

So far, we have shown that Lemma 2 holds when the SCM layers employ Gaussian signaling. The proof for discrete

signaling follows from the fact that, when the signal power tends to zero, the mutual information on a Gaussian channel is not sensitive to the input distribution. Specifically, from [15, Lemma 1], for any input distribution with zero mean, the capacity of the AWGN channel with SNR $= t$ is given by $t + o(t)$. Then, for discrete signaling, we only need to replace (A8) by

$$R_{n,i} = \frac{\Delta p_i}{p_{i+1} + r_i^{-1}} + o\left(\frac{\Delta p_i}{p_{i+1} + r_i^{-1}}\right), \quad \text{for } i = 1, \ldots, n.$$

This completes the proof of Lemma 2.

Appendix C
Proof of Theorem 1

It can be verified that $\psi(\rho) = \phi^{-1}(\rho)$ satisfies the regularity conditions required in Lemma 2. Thus, an achievable rate of the LP-LMMSE scheme is given by

$$\begin{aligned}
R &= \int_0^{+\infty} \left((\psi(\rho))^{-1} + \rho\right)^{-1} \mathrm{d}\rho \\
&= \int_{\phi(1)}^{\phi(0)} \frac{1}{(\phi^{-1}(\rho))^{-1} + \rho} \mathrm{d}\rho + \int_0^{\phi(1)} \frac{1}{1+\rho} \mathrm{d}\rho \\
&= \int_{v=1}^{v=0} \frac{1}{v^{-1} + \phi(v)} \mathrm{d}\phi(v) + \log(1 + \phi(1)) \\
&= \int_{v=1}^{v=0} \left(-\frac{\mathrm{d}\omega(v)}{\omega(v)} + \frac{\omega(v)}{v^2}\mathrm{d}v\right) - \log\omega(1) \\
&= \left(-\log\omega(v) - \frac{1}{N}\sum_{n=1}^{N} \log\left(\frac{1}{v} + \frac{d_n^2}{\sigma^2}\right)\right)\Bigg|_{v=1}^{v=0} - \log\omega(1) \\
&= \frac{1}{N}\sum_{n=1}^{N} \log\left(1 + \frac{d_n^2}{\sigma^2}\right) \qquad \text{(A11)}
\end{aligned}$$

where the first equality follows from Lemma 2, the second equality from the matching condition (24), the third equality by substituting $\rho = \phi(v)$, the fourth equality from the fact that

$$\mathrm{d}\phi(v) = -\frac{\mathrm{d}\omega(v)}{\omega(v)^2} + \frac{\mathrm{d}v}{v^2}$$

with

$$\omega(v) = \frac{1}{v^{-1} + \phi(v)} = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{v} + \frac{d_n^2}{\sigma^2}\right)^{-1},$$

and the last equality from the fact that

$$\lim_{v\to 0} \omega(v)\left(\prod_{n=1}^{N}\left(\frac{1}{v} + \frac{d_n^2}{\sigma^2}\right)\right)^{\frac{1}{N}} = 1.$$

Appendix D
Proof of Theorem 2

We first establish the relation between $mmse(\rho)$ and $\psi(\rho)$ as

$$\begin{aligned}
mmse(\rho) &= \mathrm{MMSE}(x_i' | \boldsymbol{b}', x_i' \sim p(x_i')) \\
&= \mathrm{MMSE}(x_i' | b_i', \boldsymbol{b}'_{\sim i}, x_i' \sim p(x_i')) \\
&= \mathrm{MMSE}(x_i' | b_i', a_i', x_i' \sim p(x_i')) \\
&= \gamma\left(\gamma^{-1}(\psi(\rho)) + \rho\right) \qquad \text{(A12)}
\end{aligned}$$

where the third equality follows from the fact that $b_i' \to x_i' \to \boldsymbol{x}'_{\sim i} \to \boldsymbol{b}'_{\sim i}$ forms a Markov chain (from Assumption 2 in Section II.C), together with the assumption of Theorem 2 that a_i', the extrinsic message of x_i' provided by $\boldsymbol{b}'_{\sim i}$, is modeled as an independent observation of x_i' from the effective AWGN channel with SNR $= \gamma^{-1}(\psi(\rho))$.[6] Thus

$$\begin{aligned}
R &= \int_0^{\infty} mmse(\rho)\mathrm{d}\rho \\
&= \int_0^{\infty} \gamma\left(\gamma^{-1}(\psi(\rho)) + \rho\right)\mathrm{d}\rho \\
&= \int_0^{\phi(1)} \gamma(\rho)\mathrm{d}\rho + \int_{\phi(1)}^{\phi(0)} \gamma\left(\gamma^{-1}(\phi^{-1}(\rho)) + \rho\right)\mathrm{d}\rho \\
&= \int_0^{\phi(1)} \gamma(\rho)\mathrm{d}\rho + \int_{\rho=\phi(1)}^{\rho=\phi(0)} \gamma\left(\rho' + \rho\right)\left(\mathrm{d}(\rho' + \rho) - \mathrm{d}\rho'\right) \\
&= \int_0^{\phi(1)} \gamma(\rho)\mathrm{d}\rho + \int_{\phi(1)}^{\infty} \gamma(\rho + \rho')\mathrm{d}(\rho + \rho') \\
&\quad - \int_0^{\infty} \gamma\left(\rho' + \phi(\gamma(\rho'))\right)\mathrm{d}\rho' \\
&= \log|\mathcal{S}| - \int_0^{\infty} \gamma\left(\rho + \phi(\gamma(\rho))\right)\mathrm{d}\rho
\end{aligned}$$

where the first equality follows from [14, Lemma 1], the second equality from (A12), the third equality from the curve-matching property in (24), the fourth equality by letting $\rho' = \gamma^{-1}(\phi^{-1}(\rho))$, the fifth equality by noting that ρ' varies from 0 to ∞ as ρ varies from $\phi(1)$ to $\phi(0)$, and the last equality from the equality of $\log|\mathcal{S}| = \int_0^{\infty} \gamma(\rho')\mathrm{d}\rho'$ [14].

Appendix E
Proof of Lemma 3

It can be shown that the harmonic mean

$$f(g_1, \ldots, g_N) = \left(N^{-1}\sum\nolimits_{n=1}^{N} 1/g_n\right)^{-1}$$

is concave in $\{g_n\}_{n=1}^{n=N}$ for $g_n > 0, 1 \le n \le N$. Let

$$g_n = 1/\gamma(\rho) + \lambda_n^2 w_n/\sigma^2.$$

Thus, ϕ in (15) is concave in $\{w_n\}$. Together with the fact that γ is convex and non-increasing (see [43] for the monotony of γ in ρ), $\gamma(\rho + \phi(\gamma(\rho)))$ is convex in $\{w_n\}$ according to the

[6] Here, the extrinsic variance of x_i' is $\psi(\rho) = \mathrm{MMSE}(x_i' | \boldsymbol{b}'_{\sim i}, x_i' \sim p(x_i')) = \mathrm{MMSE}(x_i' | a_i', x_i' \sim p(x_i'))$. Therefore, the SNR for the effective AWGN channel modeling a_i' is given by $\gamma^{-1}(\psi(\rho))$.

composition rule of convex functions (see [49, Ch. 3.2.4]). Nonnegative weighted sums preserve convexity. Thus

$$\int_0^{+\infty} \gamma\left(\phi(\gamma(\rho)) + \rho\right) \mathrm{d}\rho$$

is convex in $\{w_n\}$, which concludes the proof.

Appendix F
Proof of Theorem 3

For a pair of matched detector and decoder, the maximum code rate is given by

$$\begin{aligned} R &= \int_{v=1}^{v=0} \left(-\frac{\mathrm{d}\omega(v)}{\omega(v)} + \frac{\omega(v)}{v^2}\mathrm{d}v \right) - \log\omega(1) \\ &= \left(-\log\omega(v) - \frac{1}{N}\mathrm{E}\left[\log\det\left(\frac{1}{v}\boldsymbol{I} + \frac{1}{\sigma^2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^{\mathrm{H}} \right) \right] \right) \Big|_{v=1}^{v=0} \\ &\quad - \log\omega(1) \\ &= \frac{1}{N}\mathrm{E}\left[\log\det\left(\boldsymbol{I} + \frac{1}{\sigma^2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^{\mathrm{H}} \right) \right] \end{aligned}$$

where the first equality follows the derivation in Appendix C (cf., the fourth step of (A11)) except that $\omega(\lim_{v\to 0})$ is now defined as

$$\omega(v) = \frac{1}{v^{-1} + \phi(v)} = \frac{1}{N}\mathrm{tr}\left\{ \mathrm{E}\left[\left(\frac{1}{v}I + \frac{1}{\sigma^2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^{\mathrm{H}} \right)^{-1} \right] \right\},$$

and the last equality is from the fact that

$$\lim_{v\to 0} \omega(v) \left(\mathrm{E}\left[\det\left(\frac{1}{v}\boldsymbol{I} + \frac{1}{\sigma^2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^{\mathrm{H}} \right) \right] \right)^{\frac{1}{N}} = 1.$$

Acknowledgment

The authors would like to thank the anonymous reviewers for their constructive suggestions that significantly improved the presentation of this paper.

References

[1] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes. 1," in *Proc. ICC*, Geneva, Switzerland, May 1993, pp. 1064–1070.

[2] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.

[3] C. Douillard *et al.*, "Iterative correction of intersymbol interference: Turbo-equalization," *Eur. Trans. Telecommun.*, vol. 6, no. 5, pp. 507–511, Sep./Oct. 1995.

[4] 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.

[5] B. M. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, Mar. 2003.

[6] A. Sanderovich, M. Peleg, and S. Shamai, "LDPC coded MIMO multiple access with iterative joint decoding," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1437–1450, Apr. 2005.

[7] 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, Jul. 1999.

[8] J. Boutros and G. Caire, "Iterative multiuser joint decoding: Unified framework and asymptotic analysis," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 1772–1793, Jul. 2002.

[9] G. Caire, R. Muller, and T. Tanaka, "Iterative multiuser joint decoding: Optimal power allocation and low-complexity implementation," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1950–1973, Sep. 2004.

[10] X. Yuan, Q. Guo, and L. Ping, "Low-complexity iterative detection in multi-user MIMO ISI channels," *IEEE Signal Process. Lett.*, vol. 15, pp. 25–28, 2008.

[11] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 670–678, Apr. 2004.

[12] A. Ashikhmin, G. Kramer, and S. ten Brink, "Extrinsic information transfer functions: Model and erasure channel properties," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2657–2673, Nov. 2004.

[13] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.

[14] K. Bhattad and K. R. Narayanan, "An MSE-based transfer chart for analyzing iterative decoding schemes using a Gaussian approximation," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 22–38, Jan. 2007.

[15] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

[16] A. Lozano, A. M. Tulino, and S. Verdú, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3033–3051, Jul. 2006.

[17] X. Yuan, Q. Guo, X. Wang, and L. Ping, "Evolution analysis of low-cost iterative equalization in coded linear systems with cyclic prefixes," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 2, pp. 301–310, Feb. 2008.

[18] S. Verdú, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[19] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1361–1391, Jul. 1999.

[20] 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.

[21] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers. I. Unification and optimal designs," *IEEE Trans. Signal Process.*, vol. 47, no. 7, pp. 1988–2006, Jul. 1999.

[22] M. Vu and A. Paulraj, "MIMO wireless linear precoding," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 86–105, Sep. 2007.

[23] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, Sep. 2003.

[24] Y. Jiang, J. Li, and W. W. Hager, "Joint transceiver design for MIMO communications using geometric mean decomposition," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3791–3803, Oct. 2005.

[25] F. Xu, T. N. Davidson, J.-K. Zhang, and K. M. Wong, "Design of block transceivers with decision feedback detection," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 965–978, Mar. 2006.

[26] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2198–2206, Dec. 2001.

[27] D. A. Schmidt, M. Joham, and W. Utschick, "Minimum mean square error vector precoding," *Eur. Trans. Telecommun.*, vol. 19, no. 3, pp. 219–231, Apr. 2008.

[28] X. Yuan, C. Xu, L. Ping, and X. Lin, "Precoder design for multiuser MIMO ISI channels based on iterative LMMSE detection," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 6, pp. 1118–1128, Dec. 2009.

[29] F. Rey, M. Lamarca, and G. Vazquez, "Robust power allocation algorithms for MIMO OFDM systems with imperfect CSI," *IEEE Trans. Signal Process.*, vol. 53, no. 3, pp. 1070–1085, Mar. 2005.

[30] J. Wang and D. P. Palomar, "Worst-case robust MIMO transmission with imperfect channel knowledge," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3086–3100, Aug. 2009.

[31] J. Wang and D. P. Palomar, "Robust MMSE precoding in MIMO channels with pre-fixed receivers," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5802–5818, Nov. 2010.

[32] X. Yuan and L. Ping, "Space-time linear precoding and iterative LMMSE detection for MIMO channels without CSIT," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Saint Petersburg, Russia, Jul./Aug. 2011, pp. 1297–1301.

[33] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2632–2639, Sep. 2001.

[34] S. Venkatesan, S. H. Simon, and R. A. Valenzuela, "Capacity of a Gaussian MIMO channel with nonzero mean," in *Proc. IEEE Veh. Tech. Conf.*, vol. 3. Oct. 2003, pp. 1767–1771.

[35] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.

[36] R. Muller and W. H. Gerstacker, "On the capacity loss due to separation of detection and decoding," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1769–1778, Aug. 2004.

[37] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.

[38] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[39] H.-A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. R. Kschischang, "The factor graph approach to model-based signal processing," *Proc. IEEE*, vol. 95, no. 6, pp. 1295–1322, Jun. 2007.

[40] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[41] F. Pérez-Cruz, M. R. D. Rodrigues, and S. Verdú, "MIMO Gaussian channels with arbitrary inputs: Optimal precoding and power allocation," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1070–1084, Mar. 2010.

[42] M. Payaró and D. P. Palomar, "On optimal precoding in linear vector Gaussian channels with arbitrary input distribution," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun./Jul. 2009, pp. 1085–1089.

[43] D. Guo, Y. Wu, S. Shamai, and S. Verdú, "Estimation in Gaussian noise: Properties of the minimum mean-square error," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2371–2385, Apr. 2011.

[44] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 1991.

[45] K. R. Narayanan, D. N. Doan, and R. V. Tamma, "Design and analysis of LDPC codes for turbo equalization with optimal and suboptimal soft output equalizers," in *Proc. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, Oct. 2002, pp. 737–746.

[46] Y. Xin, Z. Wang, and G. B. Giannakis, "Space-time diversity systems based on linear constellation precoding," *IEEE Trans. Wireless Commun.*, vol. 2, no. 2, pp. 294–309, Mar. 2003.

[47] A. G. I. Fàbregas, A. Martinez, and G. Caire, "Bit-interleaved coded modulation," *Found. Trends Commun. Inf. Theory*, vol. 5, nos. 1–2, pp. 1–153, 2008.

[48] X. Yuan, L. Ping, and A. Kavcic, "Achievable rates of MIMO-ISI systems with linear precoding and iterative LMMSE detection," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Saint Petersburg, Russia, Jul./Aug. 2011, pp. 2909–2913.

[49] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[50] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–701, Jun. 2003.

[51] R. H. Gohary and T. N. Davidson, "On rate-optimal MIMO signalling with mean and covariance feedback," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 912–921, Feb. 2009.

[52] H. V. Poor and S. Verdú, "Probability of error in MMSE multiuser detection," *IEEE Trans. Inf. Theory*, vol. IT-43, no. 3, pp. 835–847, May 1997.

[53] Y. C. Eldar and N. Merhav, "A competitive minimax approach to robust estimation of random parameters," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1931–1946, Jul. 2004.

[54] R. B. Ash and C. Doléans-Dade, *Probability and Measure Theory*. San Francisco, CA, USA: Academic, 2000.

[55] D. Guo and S. Verdú, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1983–2010, Jun. 2005.

[56] S. Shamai and S. Verdú, "The impact of frequency-flat fading on the spectral efficiency of CDMA," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1302–1327, May 2001.

[57] L. Ping, J. Tong, X. Yuan, and Q. Guo, "Superposition coded modulation and iterative linear MMSE detection," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 6, pp. 995–1004, Aug. 2009.

[58] G. Caire and S. Shamai, "On the capacity of some channels with channel state information," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2007–2019, Sep. 1999.

Xiaojun Yuan (S'04–M'09) 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 in Electrical Engineering from the City University of Hong Kong in 2008. From 2009 to 2011, he was a research fellow at the Department of Electronic Engineering, the City University of Hong Kong. He was a visiting scholar at the Department of Electrical Engineering, the University of Hawaii at Manoa in spring and summer 2009, as well as in the same period of 2010. From 2011 to 2014, he was a research assistant professor with the Institute of Network Coding, The Chinese University of Hong Kong. He is now an assistant professor with the School of Information Science and Technology, ShanghaiTech University. His research contributions cover a broad range of wireless communications, signal processing, and information theory problems including multiple-input multiple-output (MIMO) communications, rateless codes, non-linear equalization, network coding, cooperative communications, millimeter wave communications, etc. He has published over 60 peer-reviewed research papers in the leading international journals and conferences, and has served on a number of technical programs for international conferences. He was a recipient of the Best Paper Award of IEEE International Conference on Communications in 2014.

Li Ping (S'87–M'91–SM'06–F'10) 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. He received a British Telecom-Royal Society Fellowship in 1986, the IEE J J Thomson premium in 1993, the Croucher Foundation Award in 2005 and a British Royal Academy of Engineering Distinguished Visiting Fellowship in 2010. He served as a member of Board of Governors of IEEE Information Theory Society from 2010 to 2012.

Chongbin Xu received his B.S. degree in Information Engineering from Xi'an Jiaotong University in 2005 and PhD degree in Information and Communication Engineering from Tsinghua University in 2012. He is now with the information technology group in the Department of Electronic Engineering at the City University of Hong Kong. His research interests are in the areas of signal processing and communication theory, including linear precoding, iterative detection and decoding, and cross layer design and system optimization.

Aleksandar Kavcic (SM'05) received the Dipl. Ing. degree in Electrical Engineering from Ruhr-University, Bochum, Germany in 1993, and the Ph.D. degree in Electrical and Computer Engineering from Carnegie Mellon University in 1998. Since 2007 he has been with the University of Hawaii, Honolulu where he is presently Professor of Electrical Engineering. Prior to 2007, he was in the Division of Engineering and Applied Sciences at Harvard University. He also held-short term visiting and advisory positions at City University of Hong Kong, Chinese University of Hong Kong, Seagate Technology, Read-Rite Corporation, Quantum Corporation and Link-A-Media Devices. Prof. Kavcic received the IBM Partnership Award in 1999 and the NSF CAREER Award in 2000. He is a co-recipient, with X. Ma and N. Varnica, of the 2005 IEEE Best Paper Award in Signal Processing and Coding for Data Storage. He served on the Editorial Board of the IEEE TRANSACTIONS ON INFORMATION THEORY as Associate Editor for *Detection and Estimation* from 2001 to 2004, as Guest Editor of the IEEE SIGNAL PROCESSING MAGAZINE in 2003-2004, and as Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS in 2008-2009. From 2005 until 2007, he was the Chair of the Data Storage Technical Committee of the IEEE Communications Society, and in 2014 was the general co-chair of the IEEE Symposium on Information Theory in Honolulu, Hawaii.