Linear Precoding for Finite-Alphabet Inputs Over MIMO Fading Channels With Statistical CSI

Weiliang Zeng, *Student Member, IEEE*, Chengshan Xiao, *Fellow, IEEE*, Mingxi Wang, *Student Member, IEEE*, and Jianhua Lu, *Senior Member, IEEE*

***Abstract*—This paper investigates the linear precoder design that maximizes the average mutual information of multiple-input multiple-output fading channels with statistical channel state information known at the transmitter. It formulates the design from the standpoint of finite-alphabet inputs, which leads to a problem that is very important in practice but extremely difficult in theory: First, the average mutual information lacks closed-form expression and involves prohibitive computational burden. Second, the optimization over the precoder is nonconcave and thus easily gets stuck in local maxima. To address these issues, this study first derives lower and upper bounds for the average mutual information, in which the computational complexity is reduced by several orders of magnitude compared to calculating the average mutual information directly. It proves that maximizing the bounds is asymptotically optimal and shows that, with a constant shift, the lower bound actually offers a very accurate approximation to the average mutual information for various fading channels. This paper further proposes utilizing the lower bound as a low-complexity and accurate alternative for developing a two-step algorithm to find a near global optimal precoder. Numerical examples demonstrate the convergence and efficacy of the proposed algorithm. Compared to its conventional counterparts, the proposed linear precoding method provides significant performance gain over existing precoding algorithms. The gain becomes more substantial when the spatial correlation of MIMO channels increases.**

***Index Terms*—Finite-alphabet inputs, multiple-input multiple-output, precoder design, statistical channel state information.**

I. Introduction

THE Information-theoretic approach is a fundamental means to design linear precoder. It continues to fascinate researchers, as evidenced in [1]–[5] and references therein, especially with the development of capacity achieving channel coding and transceiver structure [6]. The theoretical limit on the information rate that a communication channel can support with an arbitrary low probability of error is referred to as *channel capacity*. The capacity is achievable with independent Gaussian inputs for parallel additive white Gaussian noise (AWGN) channels and with correlated Gaussian inputs for multiple-input multiple-output (MIMO) channels [7].

Although Gaussian inputs are theoretically optimal, they are rarely realized in practice. Alternatively, in practical communication systems, inputs are usually taken from finite-alphabet constellation sets, such as phase shift keying (PSK) modulation, pulse amplitude modulation (PAM), and quadrature amplitude modulation (QAM), which depart significantly from the Gaussian assumption. Therefore, a considerable performance gap exists between precoding schemes designed based on finite-alphabet inputs and those based on Gaussian-input assumptions. In [8], the optimal power allocation for parallel Gaussian channels with finite-alphabet inputs is obtained; it demonstrates the capacity-achieving strategy for Gaussian inputs—the larger the gain of a channel, the higher its allocated power—can be quite suboptimal for finite-alphabet inputs. The reason is that the mutual information with finite-alphabet inputs is bounded; thus, there is little incentive to allocate more power to subchannels already close to saturation. For the case of MIMO channels, optimization of the precoder using a gradient-descent method is introduced in [9]–[11]. The potential computational complexity [12] and the structure of the optimal precoder [12], [13] for real-valued channels are revealed. An iterative algorithm that globally optimizes the precoder is developed in [14] for complex-valued channels.

These aforementioned results and precoding algorithms hold, unfortunately, only when the transmitter accurately knows the instantaneous channel state information (CSI). Since the dominant form of duplexing in modern wide-area mobile systems hinges on the frequency separation of the uplink and downlink, the CSI at the transmitter is obtained through feedback on bandwidth-limited control channels. In such cases, the use of instantaneous CSI is generally unrealistic, especially for fast fading channels, with two reasons: First, the overhead associated with the instantaneous feedback can be excessive. Second, the instantaneous CSI at the transmitter may be outdated when the round-trip delays obtaining the CSI are non-negligible with respect to the coherence time of channels. Therefore, exploiting long-term channel statistics appears to be more plausible because these statistics vary only with the antenna parameters and the surrounding environment and thus may change very slowly.

Although the precoding algorithms to maximize channel capacity with statistical CSI have been studied extensively for

Manuscript received April 18, 2011; revised December 19, 2011; accepted February 11, 2012. Date of publication February 22, 2012; date of current version May 11, 2012. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. David Love. This work was carried out while W. Zeng was a visiting scholar at the Missouri University of Science and Technology. This work was supported in part by the National Nature Science Foundation of China under Grant 61101071, 61021001, and 60972021 and the U.S. National Science Foundation under Grant CCF-0915846. Part of the material in this paper was presented at the IEEE Global Communications Conference (Globecom), Houston, TX, December 2011.

W. Zeng and J. Lu are with the Tsinghua National Laboratory for Information Science and Technology (TNList), Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: zengwl07@mails.tsinghua.edu.cn; lhh-dee@mail.tsinghua.edu.cn).

C. Xiao and M. Wang are with the Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO 65409 USA (e-mail: xiaoc@mst.edu; mw9zd@mail.mst.edu).

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/TSP.2012.2188717

Gaussian inputs [5], [15]–[18], it is still an open problem on how to develop an efficient and practical algorithm for maximizing the average mutual information with finite-alphabet inputs. The obstacles are twofold: First, the average mutual information lacks closed-form expression and involves prohibitive computational complexity. Second, the optimization problem over the precoding matrix is nonconcave in general and can be extremely difficult to solve [12]–[14].

This study takes a step toward exploring the structure of the optimal linear precoder that maximizes the average mutual information for finite-alphabet inputs with statistical CSI. The process begins by decomposing the precoder, with singular value decomposition (SVD), into three components: the left singular vectors, the diagonal power allocation matrix, and the right singular vectors. The left singular vectors of the optimal linear precoder are proved to be the eigenvectors of the transmit correlation matrix; the optimization of the power allocation matrix and the right singular vectors relies on iterative algorithms.

Due to the prohibitive complexity of evaluating the average mutual information, this work derives the closed-form lower and upper bounds as alternatives, which reduce the computational effort by several orders of magnitude compared to calculating the average mutual information directly. It proves that the use of either bound for precoder design achieves the optimality asymptotically in the low and high signal-to-noise ratio (SNR) regions and shows that, with a constant shift, the lower bound actually offers an accurate approximation to the average mutual information for various settings. Since the bounds are concave functions on the diagonal elements of the squared power allocation matrix, the power allocation matrix is designed within the framework of convex optimization. The optimization of the right singular vectors is formulated as a problem on the Stiefel manifold, which is solved by a gradient method with projection. Combining the design of the left singular vectors with the optimization of the power allocation matrix and the right singular vectors, a two-step iterative algorithm constitutes a unified framework, which achieves a near global optimal precoding solution.

The remainder of this paper is organized as follows: Section II introduces the system model and the properties of average mutual information. Section III considers the optimal precoding structure and introduces lower and upper bounds for average mutual information. Section IV discusses the asymptotic optimality and the concavity results, with respect to some parameters of the precoder, of both bounds. The two-step algorithm is then proposed to optimize the linear precoder, including the design of the left singular vectors, the power allocation matrix, and the right singular vectors. Section V presents several numerical examples to demonstrate the benefit for using the proposed bounds and the performance gain of the new two-step algorithm. Finally, Section VI offers conclusions.

Notation: Boldface uppercase letters denote matrices, boldface lowercase letters denote column vectors, and italics denote scalars. The superscripts $(\cdot)^T$ and $(\cdot)^H$ stand for transpose and Hermitian operations, respectively; $[\mathbf{A}]_{i,j}$ and $[\mathbf{A}]_{i,:}$ denote the (i,j)th element and ith row of matrix $\mathbf{A}$; the operator $\mathsf{diag}(\mathbf{A})$ denotes a column vector with the diagonal entries of $\mathbf{A}$, while $\mathsf{Diag}(\mathbf{a})$ denotes a diagonal matrix with elements given by vector $\mathbf{a}$; $\mathsf{Tr}(\mathbf{A})$ denotes the trace operation; $\mathbf{A} \succ 0$ denotes a positive definite matrix; $\mathbf{I}$ and $\mathbf{0}$ denote an identity matrix and a zero matrix of appropriate dimensions, respectively. The operator $\mathsf{E}_{(\cdot)}$ denotes the statistical expectation with respect to its variable, $\mathbb{C}$ denotes the complex spaces, $\Re$ is the real part of a complex number, $\jmath$ is the imaginary unit, and log and ln are used for the base two logarithm and natural logarithm, respectively.

II. System Model and Preliminaries

A. System Model

Consider a MIMO system over frequency flat fading with N_t transmit antennas and N_r receive antennas. Let $\mathbf{x} \in \mathbb{C}^{N_t \times 1}$ be a transmitted signal vector with zero mean and identity covariance matrix (i.e., $\mathsf{E}(\mathbf{x}\mathbf{x}^H) = \mathbf{I}$). The received signal $\mathbf{y} \in \mathbb{C}^{N_r \times 1}$ is given by

$$\mathbf{y} = \mathbf{H}\mathbf{P}\mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ is a random matrix whose (i,j)th entry is the complex propagation coefficient between the jth transmit antenna and the ith receive antenna; $\mathbf{P} \in \mathbb{C}^{N_t \times N_t}$ is a precoding matrix; $\mathbf{n} \in \mathbb{C}^{N_r \times 1}$ is an independent and identically distributed (i.i.d.) zero-mean circularly symmetric Gaussian noise with covariance $\sigma^2\mathbf{I}$.

For doubly correlated MIMO channels, the channel matrix $\mathbf{H}$ can be modeled as [19]

$$\mathbf{H} = \mathbf{\Psi}_{\mathbf{r}}^{\frac{1}{2}} \mathbf{H}_{\mathbf{w}} \mathbf{\Psi}_{\mathbf{t}}^{\frac{1}{2}} \tag{2}$$

where $\mathbf{H}_{\mathbf{w}} \in \mathbb{C}^{N_r \times N_t}$ is a complex matrix with i.i.d. zero-mean and unit variance Gaussian entries, and $\mathbf{\Psi}_{\mathbf{t}} \in \mathbb{C}^{N_t \times N_t} \succ 0$ and $\mathbf{\Psi}_{\mathbf{r}} \in \mathbb{C}^{N_r \times N_r} \succ 0$ are transmit and receive correlation matrices, respectively.

This work assumes the receiver has the perfect CSI through pilot-assisted channel estimation, whereas the transmitter has only statistical CSI through feedback (i.e., the transmitter knows both transmit and receive correlation matrices). Based on the statistical CSI, the precoding problem is thus the design of matrix $\mathbf{P}$ that maximizes the average mutual information.

B. Average Mutual Information for Finite-Alphabet Inputs

When input signals are drawn from the equiprobable discrete constellations, such as M-ary PSK, PAM, or QAM, where M is the number of points in the constellation sets, the average mutual information between $\mathbf{x}$ and $\mathbf{y}$, with the channel realization $\mathbf{H}$ known at the receiver, is given by

$$\mathcal{I}_A(\mathbf{P}) = \mathsf{E}_{\mathbf{H}} \mathcal{I}(\mathbf{x}; \mathbf{y}|\mathbf{H}) \tag{3}$$

in which $\mathcal{I}_A(\mathbf{P})$ is used to emphasize the dependence of average mutual information on the precoder $\mathbf{P}$; $\mathcal{I}(\mathbf{x}; \mathbf{y}|\mathbf{H})$ is the instantaneous mutual information between $\mathbf{x}$ and $\mathbf{y}$ [11]

$$\mathcal{I}(\mathbf{x}; \mathbf{y}|\mathbf{H}) = N_t \log M - \frac{1}{M^{N_t}} \sum_{m=1}^{M^{N_t}} \mathsf{E}_{\mathbf{n}} \log \sum_{k=1}^{M^{N_t}} \exp\left(-\frac{\|\mathbf{H}\mathbf{P}\mathbf{e}_{mk} + \mathbf{n}\|^2 - \|\mathbf{n}\|^2}{\sigma^2} \right) \tag{4}$$

in which $\|\cdot\|$ denotes the Euclidean norm of a vector, and $\mathbf{e}_{mk}$ is equal to $\mathbf{x}_m - \mathbf{x}_k$. Both $\mathbf{x}_m$ and $\mathbf{x}_k$ contain N_t symbols, taken independently from the M-ary signal constellation.

When a linear unitary transform $\mathbf{U}$ is applied on the channel output signal $\mathbf{y}$, the MIMO system is equivalent to a model with channel matrix $\mathbf{UH}$ and noise $\mathbf{Un}$. Considering the unitarily invariant of Euclidean norm, the average mutual information satisfies [see (3) and (4)]

$$\mathsf{E}_{\mathbf{H}}\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H}) = \mathsf{E}_{\mathbf{H}}\mathcal{I}(\mathbf{x};\mathbf{U}\mathbf{y}|\mathbf{H}). \tag{5}$$

On the other hand, when $\mathbf{U}$ is applied on $\mathbf{x}$, the average mutual information of a system with input signal $\mathbf{U}\mathbf{x}$ may change to a different value; that is,

$$\mathsf{E}_{\mathbf{H}}\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H}) \neq \mathsf{E}_{\mathbf{H}}\mathcal{I}(\mathbf{U}\mathbf{x};\mathbf{y}|\mathbf{H}) \tag{6}$$

which can be verified by numerical evaluation. This relationship implies that the linear precoder, even a unitary one, may improve the average mutual information.

The objective of this work is thus to develop efficient algorithms to find the optimal precoder that maximizes (3). The optimization is carried out over all possible $N_t \times N_t$ complex precoding matrices with transmit power constraint; it can be formulated as an optimization problem:

$$\begin{aligned} &\text{maximize} \quad \mathcal{I}_A(\mathbf{P}) \\ &\text{subject to} \quad \mathsf{Tr}\left(\mathbf{P}\mathbf{P}^H\right) \leq N_t. \end{aligned} \tag{7}$$

The computational efficiency of solving this problem depends on the calculation of $\mathcal{I}_A(\mathbf{P})$ and the optimization over $\mathbf{P}$, which are hindered by several barriers: First, the closed-form expression of $\mathcal{I}_A(\mathbf{P})$ is difficult, if not impossible, to derive, and the calculation is generally prohibitive because of the multiple integral involved. Taken $N_r \times N_t$ MIMO channels as an example, $2(N_rN_t + N_r)$ integrals from $-\infty$ to ∞ need to be considered. Second, the optimization of average mutual information can be extremely difficult because $\mathcal{I}_A(\mathbf{P})$ is nonconcave over $\mathbf{P}$. A function g is concave if its domain $\mathsf{dom}\ g$ is a convex set and if for all $\mathbf{P}_x, \mathbf{P}_y \in g$, and θ with $0 \leq \theta \leq 1$, we have [20]

$$g(\theta\mathbf{P}_x + (1-\theta)\mathbf{P}_y) \geq \theta g(\mathbf{P}_x) + (1-\theta)g(\mathbf{P}_y). \tag{8}$$

For the problem considered in this work, $\mathsf{dom}\ \mathcal{I}_A = \{\mathbf{P}|\mathsf{Tr}(\mathbf{P}\mathbf{P}^H) \leq N_t\}$ defines a convex set; however, the inequality corresponding to (8) fails to hold. This fact can be verified by counterexamples (e.g., $\mathbf{\Psi_t} = \mathbf{I}$, $\mathbf{\Psi_r} = \mathbf{I}$, $\theta = 0.5$, $\mathbf{P}_x = \mathbf{0}$, and $\mathbf{P}_y = \mathbf{I}$).

III. Optimal Precoding Structure and Average Mutual Information Bounds

This section explores the structure of the optimal precoding matrix and highlights the effect of correlation on the average mutual information so as to obtain a design criterion related only to $\mathbf{\Psi_t}$ and $\mathbf{\Psi_r}$.

From eigenvalue decomposition, the correlation matrices $\mathbf{\Psi_t}$ and $\mathbf{\Psi_r}$ can be expressed as

$$\mathbf{\Psi_t} = \mathbf{U_t}\mathbf{\Sigma_t}\mathbf{U_t}^H \quad \text{and} \quad \mathbf{\Psi_r} = \mathbf{U_r}\mathbf{\Sigma_r}\mathbf{U_r}^H \tag{9}$$

where $\mathbf{U_t}$ and $\mathbf{U_r}$ are unitary matrices whose columns are eigenvectors of $\mathbf{\Psi_t}$ and $\mathbf{\Psi_r}$, and $\mathbf{\Sigma_t}$ and $\mathbf{\Sigma_r}$ represent diagonal matrices whose diagonal entries are the eigenvalues of $\mathbf{\Psi_t}$ and $\mathbf{\Psi_r}$, respectively. Since a correlation matrix is positive-definite, the channel matrix $\mathbf{H}$ is written as

$$\mathbf{H} = \mathbf{U_r}\mathbf{\Sigma_r}^{\frac{1}{2}}\mathbf{U_r}^H\mathbf{H_w}\mathbf{U_t}\mathbf{\Sigma_t}^{\frac{1}{2}}\mathbf{U_t}^H. \tag{10}$$

Based on the property in (5) and the fact that random matrices $\mathbf{H_w}$ and $\tilde{\mathbf{H}}_\mathbf{w} = \mathbf{U_r}^H\mathbf{H_w}\mathbf{U_t}$ have the same statistics, the channel model (1) is equivalent to

$$\tilde{\mathbf{y}} = \underbrace{\mathbf{\Sigma_r}^{\frac{1}{2}}\tilde{\mathbf{H}}_\mathbf{w}\mathbf{\Sigma_t}^{\frac{1}{2}}\mathbf{U_t}^H}_{\tilde{\mathbf{H}}}\mathbf{P}\mathbf{x} + \tilde{\mathbf{n}} \tag{11}$$

where $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{n}}$ are signals when unitary transform $\mathbf{U_r}^H$ is applied on $\mathbf{y}$ and $\mathbf{n}$, respectively. Note that maximizing $\mathsf{E}_{\mathbf{H}}\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H})$ is equivalent to maximizing $\mathsf{E}_{\tilde{\mathbf{H}}_\mathbf{w}}\mathcal{I}(\mathbf{x};\tilde{\mathbf{y}}|\tilde{\mathbf{H}})$, which is the focus of the sequel discussion.

Because $\mathcal{I}(\mathbf{x};\tilde{\mathbf{y}}|\tilde{\mathbf{H}})$ is an increasing function of the SNR (i.e., $1/\sigma^2$) [10], the average mutual information $\mathsf{E}_{\tilde{\mathbf{H}}_\mathbf{w}}\mathcal{I}(\mathbf{x};\tilde{\mathbf{y}}|\tilde{\mathbf{H}})$ is also an increasing function of the SNR. Therefore, the optimal precoder should use the maximum available power; that is, the inequality constraint in problem (7) can be replaced by an equality constraint:

$$\mathsf{Tr}\left(\mathbf{P}\mathbf{P}^H\right) = N_t. \tag{12}$$

Consider SVD of the precoding matrix $\mathbf{P} = \mathbf{U_P}\mathbf{\Sigma_P}\mathbf{V_P}^H$, where $\mathbf{U_P}$ and $\mathbf{V_P}$ are unitary matrices, and $\mathbf{\Sigma_P}$ contains nonnegative diagonal entries in decreasing order, we can establish the following proposition.

Proposition 1: The left singular vectors of the optimal linear precoder that maximizes the average mutual information coincide with the eigenvectors of the transmit correlation matrix $\mathbf{\Psi_t}$.

Proof: See Appendix A. □

This result provides the design for the left singular vectors, which simplify the channel model (11) to

$$\tilde{\mathbf{y}} = \mathbf{\Sigma_r}^{\frac{1}{2}}\tilde{\mathbf{H}}_\mathbf{w}\mathbf{\Sigma_t}^{\frac{1}{2}}\tilde{\mathbf{P}}\mathbf{x} + \tilde{\mathbf{n}} \tag{13}$$

where $\tilde{\mathbf{P}} = \mathbf{\Sigma_P}\mathbf{V_P}^H$ is the remaining part of the precoder.

For the simplified model, the following proposition provides both lower and upper bounds:

Proposition 2: The average mutual information of doubly correlated MIMO channels with finite-alphabet inputs can be lower bounded by

$$\mathcal{I}_L(\tilde{\mathbf{P}}) = N_t\log M - N_r\left(\frac{1}{\ln 2} - 1\right) - \frac{1}{M^{N_t}}\sum_{m=1}^{M^{N_t}} \log\sum_{k=1}^{M^{N_t}}\prod_q\left(1 + \frac{r_q}{2\sigma^2}\mathbf{e}_{mk}^H\tilde{\mathbf{P}}^H\mathbf{\Sigma_t}\tilde{\mathbf{P}}\mathbf{e}_{mk}\right)^{-1} \tag{14}$$

where r_q denotes the qth diagonal element of $\mathbf{\Sigma_r}$; it can be upper bounded by

$$\mathcal{I}_U(\tilde{\mathbf{P}}) = N_t\log M - \frac{1}{M^{N_t}}\sum_{m=1}^{M^{N_t}} \log\sum_{k=1}^{M^{N_t}}\exp\left(-\frac{\mathsf{Tr}(\mathbf{\Sigma_r})}{\sigma^2}\mathbf{e}_{mk}^H\tilde{\mathbf{P}}^H\mathbf{\Sigma_t}\tilde{\mathbf{P}}\mathbf{e}_{mk}\right). \tag{15}$$

Proof: See Appendix B. □

IV. Precoder Design for Maximizing the Average Mutual Information

This section considers the precoder design with criteria of maximizing either the lower bound (14) or the upper bound (15). It begins by proving the asymptotic optimality of using both bounds and then develops a unified algorithm to solve the precoding problem efficiently.

A. Asymptotic Optimality

1) Asymptotic Optimality at Low SNR Region: According to (14), the problem of maximizing $\mathcal{I}_L(\tilde{\mathbf{P}})$ is equivalent to the following minimization problem:

$$\begin{aligned} \text{minimize} \quad & \sum_{m=1}^{M^{N_t}} \log \sum_{k=1}^{M^{N_t}} \prod_q \left(1+\frac{r_q}{2\sigma^2}\mathbf{e}_{mk}^H \tilde{\mathbf{P}}^H \boldsymbol{\Sigma}_{\mathbf{t}} \tilde{\mathbf{P}} \mathbf{e}_{mk}\right)^{-1} \\ \text{subject to} \quad & \mathsf{Tr}\left(\tilde{\mathbf{P}}\tilde{\mathbf{P}}^H\right) = N_t. \end{aligned} \tag{16}$$

When $\sigma^2 \to +\infty$ (i.e., at low SNR region), the objective function in (16) can be expressed, based on Taylor expansion, as

$$\begin{aligned} &\sum_{m=1}^{M^{N_t}} \log \sum_{k=1}^{M^{N_t}} \prod_q \left(1+\frac{r_q}{2\sigma^2}\mathbf{e}_{mk}^H \tilde{\mathbf{P}}^H \boldsymbol{\Sigma}_{\mathbf{t}} \tilde{\mathbf{P}} \mathbf{e}_{mk}\right)^{-1} \\ &\quad = M^{N_t}\log M^{N_t} - \frac{\mathsf{Tr}(\boldsymbol{\Sigma}_{\mathbf{r}})}{2\ln(2) M^{N_t}\sigma^2} \\ &\qquad \times \left(\sum_m^{M^{N_t}} \sum_k^{M^{N_t}} \mathbf{e}_{mk}^H \tilde{\mathbf{P}}^H \boldsymbol{\Sigma}_{\mathbf{t}} \tilde{\mathbf{P}} \mathbf{e}_{mk}\right) + \mathcal{O}\left(\frac{1}{\sigma^4}\right) \end{aligned} \tag{17}$$

where $\mathcal{O}\left(\frac{1}{\sigma^4}\right)$ denotes the least significant terms on the order of $\frac{1}{\sigma^4}$. The solution of maximizing $\mathcal{I}_L(\tilde{\mathbf{P}})$ for first-order optimality is thus equivalent to the problem

$$\begin{aligned} \text{maximize} \quad & \sum_m^{M^{N_t}} \sum_k^{M^{N_t}} \mathbf{e}_{mk}^H \tilde{\mathbf{P}}^H \boldsymbol{\Sigma}_{\mathbf{t}} \tilde{\mathbf{P}} \mathbf{e}_{mk} \\ \text{subject to} \quad & \mathsf{Tr}\left(\tilde{\mathbf{P}}\tilde{\mathbf{P}}^H\right) = N_t. \end{aligned} \tag{18}$$

For a equiprobable zero-mean constellation set, the objective function can be reformulated as

$$\begin{aligned} &\sum_m^{M^{N_t}} \sum_k^{M^{N_t}} \mathbf{e}_{mk}^H \tilde{\mathbf{P}}^H \boldsymbol{\Sigma}_{\mathbf{t}} \tilde{\mathbf{P}} \mathbf{e}_{mk} \\ &\quad = \sum_m^{M^{N_t}} \sum_k^{M^{N_t}} \mathsf{Tr}\left(\tilde{\mathbf{P}}^H \boldsymbol{\Sigma}_{\mathbf{t}} \tilde{\mathbf{P}} \mathbf{e}_{mk}\mathbf{e}_{mk}^H\right) = e_{mk}\mathsf{Tr}\left(\boldsymbol{\Sigma}_{\mathbf{t}}\boldsymbol{\Sigma}_{\mathbf{P}}^2\right) \end{aligned} \tag{19}$$

where e_{mk} is a constant with $e_{mk}\mathbf{I} = \sum_m^{M^{N_t}} \sum_k^{M^{N_t}} \mathbf{e}_{mk}\mathbf{e}_{mk}^H$. The problem of (18) is thus equivalent to

$$\begin{aligned} \text{maximize} \quad & \mathsf{Tr}\left(\boldsymbol{\Sigma}_{\mathbf{t}}\boldsymbol{\Sigma}_{\mathbf{P}}^2\right) \\ \text{subject to} \quad & \mathsf{Tr}\left(\boldsymbol{\Sigma}_{\mathbf{P}}^2\right) = N_t \end{aligned} \tag{20}$$

which is solved when the diagonal coefficients of $\boldsymbol{\Sigma}_{\mathbf{P}}^2$ are zero except for elements corresponding to the maximum eigenvalues of $\boldsymbol{\Psi}_{\mathbf{t}}$. Combining the optimal precoding structure in Proposition 1, the following proposition holds:

Proposition 3: The optimal linear precoder to maximize the lower bound of the average mutual information at low SNR region is equal to $\mathbf{U}_{\mathbf{t}}$ times a diagonal power allocation matrix with all power allocated on the maximum eigenvalues of $\boldsymbol{\Psi}_{\mathbf{t}}$ (i.e., the beamforming strategy). □

At the same time, precoder maximizes the average mutual information directly at lower SNR region can be solved for first-order optimality by [10]

$$\begin{aligned} \text{maximize} \quad & \mathsf{E}_{\mathbf{H}}\mathsf{Tr}\left(\mathbf{P}^H\mathbf{H}^H\mathbf{H}\mathbf{P}\right) \\ \text{subject to} \quad & \mathsf{Tr}\left(\mathbf{P}^H\mathbf{P}\right) = N_t \end{aligned} \tag{21}$$

in which the objective function can be rewritten as

$$\begin{aligned} &\mathsf{E}_{\mathbf{H}}\mathsf{Tr}\left(\mathbf{P}^H\mathbf{H}^H\mathbf{H}\mathbf{P}\right) \\ &\quad = \mathsf{E}_{\tilde{\mathbf{H}}_{\mathbf{w}}}\mathsf{Tr}\left(\mathbf{P}^H\mathbf{U}_{\mathbf{t}}\boldsymbol{\Sigma}_{\mathbf{t}}^{\frac{1}{2}}\tilde{\mathbf{H}}_{\mathbf{w}}^H\boldsymbol{\Sigma}_{\mathbf{r}}\tilde{\mathbf{H}}_{\mathbf{w}}\boldsymbol{\Sigma}_{\mathbf{t}}^{\frac{1}{2}}\mathbf{U}_{\mathbf{t}}^H\mathbf{P}\right) \\ &\quad = \mathsf{Tr}(\boldsymbol{\Sigma}_{\mathbf{r}})\cdot\mathsf{Tr}\left(\mathbf{P}^H\mathbf{U}_{\mathbf{t}}\boldsymbol{\Sigma}_{\mathbf{t}}\mathbf{U}_{\mathbf{t}}^H\mathbf{P}\right) \\ &\quad = \mathsf{Tr}(\boldsymbol{\Sigma}_{\mathbf{r}})\cdot\mathsf{Tr}\left(\boldsymbol{\Sigma}_{\mathbf{P}}^2\mathbf{U}_{\mathbf{P}}^H\mathbf{U}_{\mathbf{t}}\boldsymbol{\Sigma}_{\mathbf{t}}\mathbf{U}_{\mathbf{t}}^H\mathbf{U}_{\mathbf{P}}\right). \end{aligned} \tag{22}$$

This function is maximized when $\mathbf{U}_{\mathbf{P}}$ is equal to $\mathbf{U}_{\mathbf{t}}$ and the coefficients of $\boldsymbol{\Sigma}_{\mathbf{P}}^2$ are zeros except for the elements corresponding to the maximum eigenvalues of $\boldsymbol{\Psi}_{\mathbf{t}}$. This solution is the same as that of maximizing the lower bound at low SNR region; see Proposition 3. That is, maximizing the lower bound is asymptotically optimal compared with maximizing the average mutual information directly.

Following the similar steps, the asymptotic optimality of maximizing the upper bound at low SNR region can be proved, and it is omitted here for brevity.

2) Asymptotic Optimality at High SNR Region: The proof of asymptotic optimality of maximizing the lower bound at high SNR region is offered here. Similar steps can also be followed to maximize the upper bound. The objective function in (16) is first rewritten as

$$\begin{aligned} &\sum_{m=1}^{M^{N_t}} \log \sum_{k=1}^{M^{N_t}} \prod_q \left(1+\frac{r_q}{2\sigma^2}\mathbf{e}_{mk}^H \tilde{\mathbf{P}}^H \boldsymbol{\Sigma}_{\mathbf{t}} \tilde{\mathbf{P}} \mathbf{e}_{mk}\right)^{-1} \\ &\quad = \sum_{m=1}^{M^{N_t}} \log \sum_{k=1}^{M^{N_t}} \\ &\qquad \exp\left[\ln \prod_q \left(1+\frac{r_q}{2\sigma^2}\mathbf{e}_{mk}^H \tilde{\mathbf{P}}^H \boldsymbol{\Sigma}_{\mathbf{t}} \tilde{\mathbf{P}} \mathbf{e}_{mk}\right)^{-1}\right]. \end{aligned} \tag{23}$$

Note that $\log\sum_{k=1}^{M^{N_t}} \exp(\cdot)$ is a soft version of maximization [20], and it is dominated—with the help of exponential operator—by the most significant term. The idea here is to replace the soft maximization by its hard version:

$$\begin{aligned} &\sum_{m=1}^{M^{N_t}} \log \sum_{k=1}^{M^{N_t}} \exp\left[\ln \prod_q \left(1+\frac{r_q}{2\sigma^2}\mathbf{e}_{mk}^H \tilde{\mathbf{P}}^H \boldsymbol{\Sigma}_{\mathbf{t}} \tilde{\mathbf{P}} \mathbf{e}_{mk}\right)^{-1}\right] \\ &\quad \approx \sum_{m=1}^{M^{N_t}} \max_k \ln \prod_q \left(1+\frac{r_q}{2\sigma^2}\mathbf{e}_{mk}^H \tilde{\mathbf{P}}^H \boldsymbol{\Sigma}_{\mathbf{t}} \tilde{\mathbf{P}} \mathbf{e}_{mk}\right)^{-1} \\ &\quad = \sum_{m=1}^{M^{N_t}} \max_k \left[-\sum_q \ln\left(1+\frac{r_q}{2\sigma^2}\mathbf{e}_{mk}^H \tilde{\mathbf{P}}^H \boldsymbol{\Sigma}_{\mathbf{t}} \tilde{\mathbf{P}} \mathbf{e}_{mk}\right)\right]. \end{aligned}$$

TABLE I
ASYMPTOTIC RESULTS OF OPTIMAL LINEAR PRECODING

Cases	Optimal Precoder Design	
	At Low SNR Region	At High SNR Region
Maximize instantaneous mutual information with full CSI at the transmitter	Beamforming (See [10] for reference)	Maximize minimum distance (See [10] for reference)
Maximize lower bound function (14) with statistical CSI at the transmitter	Beamforming (Derived in Sec. IV-A)	Maximize minimum distance (Derived in Sec. IV-A)
Maximize upper bound function (15) with statistical CSI at the transmitter	Beamforming (Derived in Sec. IV-A)	Maximize minimum distance (Derived in Sec. IV-A)

When $\sigma^2 \rightarrow 0$ (i.e., at high SNR region), it yields

$$\begin{aligned}
&\sum_{m=1}^{M^{N_t}} \max_k \left[-\sum_q \ln\left(1+\frac{r_q}{2\sigma^2}\mathbf{e}_{mk}^H \tilde{\mathbf{P}}^H \boldsymbol{\Sigma}_{\mathbf{t}} \tilde{\mathbf{P}} \mathbf{e}_{mk}\right)\right] \\
&\quad \approx \sum_{m=1}^{M^{N_t}} \max_k \left[-\sum_q \ln\left(\frac{r_q}{2\sigma^2}\mathbf{e}_{mk}^H \tilde{\mathbf{P}}^H \boldsymbol{\Sigma}_{\mathbf{t}} \tilde{\mathbf{P}} \mathbf{e}_{mk}\right)\right] \\
&\quad \leq -\min_{\substack{m,k \\ m\neq k}} \left[\sum_q \ln\left(\frac{r_q}{2\sigma^2}\right) + N_r \ln\left(\mathbf{e}_{mk}^H \tilde{\mathbf{P}}^H \boldsymbol{\Sigma}_{\mathbf{t}} \tilde{\mathbf{P}} \mathbf{e}_{mk}\right)\right].
\end{aligned}$$

Therefore, the optimization problem in (16) corresponds, at high SNR region, to

$$\begin{aligned}
\text{maximize} \quad & \min_{\substack{m,k \\ m\neq k}} \mathbf{e}_{mk}^H \tilde{\mathbf{P}}^H \boldsymbol{\Sigma}_{\mathbf{t}} \tilde{\mathbf{P}} \mathbf{e}_{mk} \\
\text{subject to} \quad & \mathsf{Tr}\left(\tilde{\mathbf{P}}\tilde{\mathbf{P}}^H\right) = N_t. \qquad (24)
\end{aligned}$$

The minimization of the quadratic form in (24) can be identified as the minimum distance among all possible realizations of the input vector

$$d_{\min} = \min_{\substack{m,k \\ m\neq k}} \|\boldsymbol{\Sigma}_{\mathbf{t}}^{\frac{1}{2}} \tilde{\mathbf{P}}(\mathbf{x}_m - \mathbf{x}_k)\|^2 \qquad (25)$$

which leads to the following results:

Proposition 4: The optimal linear precoder to maximize the lower bound of the average mutual information at high SNR region is equivalent to maximizing the minimum distance between the constellation vectors. □

At the same time, precoder design that maximizes the average mutual information directly at high SNR region suggests to maximize the average minimum distance among all possible realizations of the input vector

$$d_{\min}^{(A)} = \min_{\substack{m,k \\ m\neq k}} \mathsf{E}_{\tilde{\mathbf{H}}_{\mathbf{w}}} \|\boldsymbol{\Sigma}_{\mathbf{r}}^{\frac{1}{2}} \tilde{\mathbf{H}}_{\mathbf{w}} \boldsymbol{\Sigma}_{\mathbf{t}}^{\frac{1}{2}} \tilde{\mathbf{P}}(\mathbf{x}_m - \mathbf{x}_k)\|^2. \qquad (26)$$

The above criterion is based on the result of instantaneous mutual information with full CSI at the transmitter [10] and the simplified channel model in (13). Since $\tilde{\mathbf{H}}_{\mathbf{w}}$ is a complex random matrix with zero-mean and unit variance Gaussian entries, the expectation over $\tilde{\mathbf{H}}_{\mathbf{w}}$ in (26) can be derived as

$$d_{\min}^{(A)} = \min_{\substack{m,k \\ m\neq k}} \mathsf{Tr}(\boldsymbol{\Sigma}_{\mathbf{r}}) \cdot \|\boldsymbol{\Sigma}_{\mathbf{t}}^{\frac{1}{2}} \tilde{\mathbf{P}}(\mathbf{x}_m - \mathbf{x}_k)\|^2. \qquad (27)$$

Thus, the criterion of maximizing the average mutual information directly leads to the same precoder as that of maximizing the lower bound; that is, it is asymptotically optimal to maximize the lower bound at high SNR region. Table I summarizes the asymptotic results for the cases considered in this section.

B. Concavity Results

Due to the low computational complexity and asymptotic optimality of the lower and upper bounds, it is desirable to employ them for maximizing the average mutual information. The development of an efficient algorithm to solve the problem depends on concavity property, which leads to global optimality, with respect to design parameters.

The first option is to identify the concavity of $\tilde{\mathbf{P}}$. However, neither the lower bound nor the upper bound is a concave function over $\tilde{\mathbf{P}}$. This fact can be verified by a specific counterexample (e.g., $\mathbf{H} \in \mathbb{C}^{1\times 1}$).

The second option is to identify the concavity of $\tilde{\mathbf{P}}^H \boldsymbol{\Sigma}_{\mathbf{t}} \tilde{\mathbf{P}}$. Similar form has been stated to be concave for real-valued channels with instantaneous CSI at the transmitter in [13] and has been rigorously proved for complex-valued channels in [14]. The following proposition shows this result also holds for both the lower and upper bounds with statistical CSI.

Proposition 5: Both the lower and upper bounds of the average mutual information, derived in (14) and (15), are concave functions of $\tilde{\mathbf{P}}^H \boldsymbol{\Sigma}_{\mathbf{t}} \tilde{\mathbf{P}}$.

Proof: See Appendix C. □

Although the concavity results in Proposition 5 hold, it is difficult to transform the power constraint $\mathsf{Tr}(\tilde{\mathbf{P}}\tilde{\mathbf{P}}^H) = N_t$ to an equivalent convex constraint over $\tilde{\mathbf{P}}^H \boldsymbol{\Sigma}_{\mathbf{t}} \tilde{\mathbf{P}}$. The concavity over the squared singular values of $\mathbf{P}$ is then identified as another option:

Proposition 6: Both the lower and upper bounds of the average mutual information are concave functions of the power allocation vector $\boldsymbol{\lambda} = \mathsf{diag}(\boldsymbol{\Sigma}_{\mathbf{P}}^2)$. The gradient of the lower bound with respect to $\boldsymbol{\lambda}$ is given by

$$\nabla_{\boldsymbol{\lambda}} \mathcal{I}_L = \mathsf{diag}\left(\sum_{m=1}^{M^{N_t}} \frac{\sum_{k=1}^{M^{N_t}} \left\{ w_{mk} \cdot \sum_q \frac{r_q \mathbf{W}_{mk}^T}{2\sigma^2 + r_q \mathsf{Tr}(\mathbf{W}_{mk}\boldsymbol{\Sigma}_{\mathbf{P}}^2)} \right\}}{\ln(2) M^{N_t} \cdot \sum_{k=1}^{M^{N_t}} w_{mk}} \right) \qquad (28)$$

with

$$\mathbf{W}_{mk} = \boldsymbol{\Sigma}_{\mathbf{t}} \mathbf{V}_{\mathbf{P}}^H \mathbf{e}_{mk} \mathbf{e}_{mk}^H \mathbf{V}_{\mathbf{P}} \qquad (29)$$

and

$$w_{mk} = \exp\left[-\sum_q \ln\left(1 + \frac{r_q}{2\sigma^2} \mathsf{Tr}\left(\mathbf{W}_{mk}\boldsymbol{\Sigma}_{\mathbf{P}}^2\right)\right)\right]. \qquad (30)$$

TABLE II
A SUMMARY OF CONCAVITY AND NONCONCAVITY RESULTS FOR VARIOUS CASES. A $\surd$ INDICATES THAT THE CORRESPONDING CONCAVITY RESULT HOLDS; AN $\times$ INDICATES THAT IT DOES NOT

Cases	Concavity over $\tilde{\mathbf{P}}$	Concavity over $\tilde{\mathbf{P}}^H\mathbf{\Sigma_t}\tilde{\mathbf{P}}$	Concavity over $\boldsymbol{\lambda} = \mathsf{diag}(\mathbf{\Sigma}_{\mathbf{P}}^2)$
Instantaneous mutual information with full CSI at the transmitter	$\times$ (See [14] and [21])	$\surd$ (See [13] and [14])	$\surd$ (See [12], [14], and [21])
Lower bound function (14) with statistical CSI at the transmitter	$\times$ (See Sec. IV-B)	$\surd$ (Proved in Proposition 5)	$\surd$ (Proved in Proposition 6)
Upper bound function (15) with statistical CSI at the transmitter	$\times$ (See Sec. IV-B)	$\surd$ (Proved in Proposition 5)	$\surd$ (Proved in Proposition 6)

At the same time, the gradient of upper bound with respect to $\boldsymbol{\lambda}$ is given by

$$\nabla_{\boldsymbol{\lambda}}\mathcal{I}_U = \mathsf{diag}\left(\frac{\mathsf{Tr}(\mathbf{\Sigma_r})}{M^{N_t}}\sum_{m=1}^{M^{N_t}}\frac{\sum_{k=1}^{M^{N_t}}\left\{f_{mk}\cdot\mathbf{W}_{mk}^T\right\}}{\ln(2)\sigma^2\cdot\sum_{k=1}^{M^{N_t}}f_{mk}}\right) \tag{31}$$

with

$$f_{mk} = \exp\left[-\frac{\mathsf{Tr}(\mathbf{\Sigma_r})}{\sigma^2}\mathsf{Tr}(\mathbf{W}_{mk}\mathbf{\Sigma}_{\mathbf{P}}^2)\right]. \tag{32}$$

Proof: See Appendix C. □

Table II summarizes the concavity and nonconcavity results for the lower and upper bounds. For comparison, it also provides the corresponding results for instantaneous mutual information with full CSI at the transmitter.

C. Precoder Design

This section provides a unified precoder design framework and takes the precoder design problem of maximizing the lower bound as an example. The approach proposed here can also be used to maximize the upper bound.

1) Optimal Power Allocation: The concavity result in Proposition 6 ensures that a global optimal power allocation vector can be found given a right singular vectors $\mathbf{V_P}$. The first step is thus to optimize the power allocation vector $\boldsymbol{\lambda}$ given feasible right singular vectors:

$$\begin{aligned}\text{maximize}\quad & \mathcal{I}_L(\boldsymbol{\lambda})\\ \text{subject to}\quad & \mathbf{1}^T\boldsymbol{\lambda} = N_t\\ & \boldsymbol{\lambda}\succeq\mathbf{0}\end{aligned} \tag{33}$$

where $\mathbf{1}$ denotes the column vector with all entries one.

To solve this problem efficiently, the sequel exploits the structure of the problem and develops an iterative algorithm to achieve the global optimum.

First, problem (33) is re-expressed, making the inequality constraints implicit in the objective function using the *barrier* method [20]:

$$\begin{aligned}\text{minimize}\quad & f(\boldsymbol{\lambda}) = -\mathcal{I}_L(\boldsymbol{\lambda}) + \sum_{i=1}^{N_t}\phi(-\lambda_i)\\ \text{subject to}\quad & \mathbf{1}^T\boldsymbol{\lambda} = N_t\end{aligned} \tag{34}$$

where $\phi(u)$ is the logarithmic barrier function, approximating an indicator of whether constraints are violated:

$$\phi(u) = \begin{cases}-\left(\frac{1}{t}\right)\ln(-u), & u<0\\ +\infty, & u\geq 0.\end{cases} \tag{35}$$

The parameter $t>0$ sets the accuracy of the approximation. As it grows, the quality of the approximation improves.

The next step is to address the equality constraint. At the feasible point $\boldsymbol{\lambda}$, $f(\boldsymbol{\lambda})$ can be approximated near $\boldsymbol{\lambda}$ by its linear expansion:

$$f(\boldsymbol{\lambda}+\mathbf{d}) \approx f(\boldsymbol{\lambda}) + \nabla_{\boldsymbol{\lambda}}f(\boldsymbol{\lambda})^T\mathbf{d} \tag{36}$$

where $\nabla_{\boldsymbol{\lambda}}f(\boldsymbol{\lambda})$ is the gradient of f at $\boldsymbol{\lambda}$,

$$\nabla_{\boldsymbol{\lambda}}f(\boldsymbol{\lambda}) = -\nabla_{\boldsymbol{\lambda}}\mathcal{I}_L(\boldsymbol{\lambda}) - \left(\frac{1}{t}\right)\mathbf{q} \tag{37}$$

with $\nabla_{\boldsymbol{\lambda}}\mathcal{I}_L(\boldsymbol{\lambda})$ provided in (28) and $q_i = \frac{1}{\lambda_i}$ is the ith element of vector $\mathbf{q}$. This approximation leads to the following direction finding problem in order to solve (34):

$$\begin{aligned}\text{minimize}\quad & \nabla_{\boldsymbol{\lambda}}f(\boldsymbol{\lambda})^T\mathbf{d}\\ \text{subject to}\quad & \mathbf{1}^T(\boldsymbol{\lambda}+\mathbf{d}) = N_t\\ & \mathbf{d}^T\mathbf{d}\leq 1\end{aligned} \tag{38}$$

where the first constraint in (38) ensures that the equality constraint of the original problem is satisfied, and the second constraint restricts the direction $\mathbf{d}$ in the Euclidean unit ball.

This minimization problem can be solved analytically. We define $\Delta\boldsymbol{\lambda}$, the descent direction, as the scaled solution of problem (38), which is characterized by the following Karush–Kuhn–Tucker (KKT) systems:

$$\begin{bmatrix}\mathbf{I} & \mathbf{1}\\ \mathbf{1}^T & 0\end{bmatrix}\begin{bmatrix}\Delta\boldsymbol{\lambda}\\ \nu\end{bmatrix} = \begin{bmatrix}-\nabla_{\boldsymbol{\lambda}}f(\boldsymbol{\lambda})\\ 0\end{bmatrix} \tag{39}$$

where ν is the Lagrange multiplier associated with the equality constraint. Considering the special structure of this linear system and applying block matrix inversion, $\Delta\boldsymbol{\lambda}$ is solved by

$$\Delta\boldsymbol{\lambda} = -\left(\mathbf{I} - \frac{\mathbf{1}\cdot\mathbf{1}^T}{N_t}\right)\nabla_{\boldsymbol{\lambda}}f(\boldsymbol{\lambda}) \tag{40}$$

where $\left(\mathbf{I}-\mathbf{1}\cdot\mathbf{1}^T/N_t\right)$ is a matrix, by which the gradient of f at $\boldsymbol{\lambda}$, or $\nabla_{\boldsymbol{\lambda}}f(\boldsymbol{\lambda})$, is projected onto the space that constrains the search direction to

$$\mathbf{1}^T\Delta\boldsymbol{\lambda} = 0. \tag{41}$$

Combining this search direction with the backtracking line search conditions [20], Algorithm 1, which is presented below, finds the optimal power allocation vector. It ensures convergence to the global optimal power allocation vector because of the concavity.

Algorithm 1: Maximizing the Lower Bound of Average Mutual Information Over the Power Allocation Vector:

1. *Initialization*. Given a feasible vector $\boldsymbol{\lambda}$, $t := t^{(0)} > 0$, $\alpha > 1$, and tolerance $\epsilon > 0$.
2. *Search direction*. Compute the gradient of f at $\boldsymbol{\lambda}$, $\nabla_{\boldsymbol{\lambda}} f(\boldsymbol{\lambda})$, as (37) and the descent direction $\Delta\boldsymbol{\lambda}$ as (40). Evaluate $\|\Delta\boldsymbol{\lambda}\|^2$. If it is sufficiently small, then go to Step 5.
3. *Search step*. Choose step size γ by backtracking line search.
4. *Update*. Set $\boldsymbol{\lambda} := \boldsymbol{\lambda} + \gamma\Delta\boldsymbol{\lambda}$. Go to Step 2.
5. *Iteration and stop*. Stop if $\frac{1}{t} < \epsilon$, else $t := \alpha t$, and go to step 2.

2) Optimization Over Right Singular Vectors: Given a power allocation vector $\boldsymbol{\lambda}$, the maximization of the lower bound on the linear precoder becomes the following:

$$\begin{aligned} &\text{maximize} && \mathcal{I}_L(\mathbf{V_P}) \\ &\text{subject to} && \mathbf{V_P}^H\mathbf{V_P} = \mathbf{I} \end{aligned} \tag{42}$$

where $\mathcal{I}_L(\mathbf{V_P})$ emphasizes the dependent of $\mathcal{I}_L$ on $\mathbf{V_P}$.

With unitary matrix constraint, (42) can be reformulated as an unconstrained problem in a constrained search space:

$$\text{minimize} \quad h(\mathbf{V_P}) \tag{43}$$

where $h(\mathbf{V_P})$ is defined as $-\mathcal{I}_L(\mathbf{V_P})$ with its domain restricted to the complex Stiefel manifold:

$$\mathsf{dom}\ h = \left\{\mathbf{V_P} \in \mathbb{S}^{N_t\times N_t}\right\} \tag{44}$$

and

$$\mathbb{S}^{N_t\times N_t} = \left\{\mathbf{U} \in \mathbb{C}^{N_t\times N_t} | \mathbf{U}^H\mathbf{U} = \mathbf{I}\right\}. \tag{45}$$

To solve this unitary-matrix-constrained problem, the negative gradient of h on tangent space is suggested as the descent direction [22]:

$$\Delta\mathbf{V_P} = \nabla_{\mathbf{V_P}}\mathcal{I}_L(\mathbf{V_P}) - \mathbf{V_P}\left(\nabla_{\mathbf{V_P}}\mathcal{I}_L(\mathbf{V_P})\right)^H\mathbf{V_P} \tag{46}$$

where $\nabla_{\mathbf{V_P}}\mathcal{I}_L(\mathbf{V_P})$ is the gradient of the lower bound. This gradient is given by

$$\nabla_{\mathbf{V_P}}\mathcal{I}_L(\mathbf{V_P}) = \sum_{m=1}^{M^{N_t}} \frac{\left\{\sum_{k=1}^{M^{N_t}} w_{mk}\cdot\sum_q \frac{r_q\mathbf{Q}_{mk}}{2\sigma^2+r_q\mathsf{Tr}(\mathbf{V_P}^H\mathbf{Q}_{mk})}\right\}}{\ln(2)M^{N_t}\cdot\sum_{k=1}^{M^{N_t}} w_{mk}} \tag{47}$$

with w_{mk} defined in (30) and $\mathbf{Q}_{mk}$ given by

$$\mathbf{Q}_{mk} = \mathbf{e}_{mk}\mathbf{e}_{mk}^H\mathbf{V_P}\boldsymbol{\Sigma}_{\mathbf{P}}^2\boldsymbol{\Sigma}_{\mathbf{t}}. \tag{48}$$

However, movement on the tangent space in the direction of descent may fail to maintain the variable on the Stiefel manifold; therefore, the unitary property must be restored at each iteration step by projecting an infeasible point onto $\mathbb{S}^{N_t\times N_t}$. The projection of an arbitrary matrix $\mathbf{W} \in \mathbb{C}^{n\times n}$ is defined as the point on the Stiefel manifold $\pi(\mathbf{W})$ closest to $\mathbf{W}$ in the Euclidean norm:

$$\pi(\mathbf{W}) = \arg\min_{\mathbf{Q}\in\mathbb{S}^{N_t\times N_t}} \|\mathbf{W}-\mathbf{Q}\|^2. \tag{49}$$

Proposition 7 (Projection): Given a full-rank matrix $\mathbf{W} \in \mathbb{C}^{n\times n}$ with SVD defined as $\mathbf{W} = \mathbf{U_W}\boldsymbol{\Sigma}\mathbf{V_W}^H$, the projection is unique, as expressed by $\pi(\mathbf{W}) = \mathbf{U_W}\mathbf{V_W}^H$ [23, Sec. 7.4.8].

By combining the search direction (46) and the projection with the backtracking line search, an optimization algorithm is developed to maximize the lower bound over the right singular vectors $\mathbf{V_P}$.

Algorithm 2: Maximizing the Lower Bound of Average Mutual Information Over Right Singular Vectors:

1. *Initialization*. Given a feasible $\mathbf{V_P} \in \mathbb{S}^{n_t\times N_t}$.
2. *Search direction*. Compute the Descent Direction of h at $\mathbf{V_P}$, $\delta\mathbf{V_P}$, as (46). Evaluate $\|\Delta\mathbf{V_P}\|^2 = \mathsf{Tr}\{(\Delta\mathbf{V_P})^H\Delta\mathbf{V_P}\}$. If it is sufficiently small, then stop; else, go to Step 3.
3. *Search step*. Choose step size γ as follows:
   a) If $h(\mathbf{V_P}) - h(\pi(\mathbf{V_P}+2\gamma\Delta\mathbf{V_P})) \geq \gamma\|\Delta\mathbf{V_P}\|^2$, then set $\gamma := 2\gamma$, and repeat Step 3a).
   b) If $h(\mathbf{V_P}) - h(\pi(\mathbf{V_P}+\gamma\Delta\mathbf{V_P})) < \frac{1}{2}\gamma\|\Delta\mathbf{V_P}\|^2$, then set $\gamma := \frac{1}{2}\gamma$, and repeat Step 3b).
4. *Update*. Set $\mathbf{V_P} := \pi(\mathbf{V_P} + \gamma\Delta\mathbf{V_P})$. Go to Step 2.

3) Two-Step Approach to Optimize Precoder: A complete two-step approach can now be developed by combining the design for left singular vectors in Proposition 1, the design for the power allocation vector in Algorithm 1, and the design for the right singular vectors in Algorithm 2.

Algorithm 3: Two-Step Algorithm to Maximize the Lower Bound Over a Linear Precoder:

1. *Initialization*. Given feasible initial points $\boldsymbol{\lambda}^{(0)}$ and $\mathbf{V_P}^{(0)}$, and set $n := 1$.
2. *Left singular vectors*. Let $\mathbf{U_P}$ equal the eigenvectors of transmit correlation matrix, $\mathbf{U_t}$.
3. *Update power allocation vector*. Solve the following problem by Algorithm 1:
$$\boldsymbol{\lambda}^{(n)} := \arg\max_{\substack{\mathbf{1}^T\boldsymbol{\lambda}=N_t \\ \boldsymbol{\lambda}\succeq\mathbf{0}}} \mathcal{I}_L(\boldsymbol{\lambda}, \mathbf{V_P}^{(n-1)}).$$
4. *Update right singular vectors*. Solve the following problem by Algorithm 2:
$$\mathbf{V_P}^{(n)} := \arg\max_{\mathbf{V_P}\in\mathbb{S}^{N_t\times N_t}} \mathcal{I}_L(\boldsymbol{\lambda}^{(n)}, \mathbf{V_P}).$$
5. *Iteration*. Set $n := n+1$. Go to Step 3 until convergence.
6. *Output*. The optimal precoder is thus given by $\mathbf{P}^\star = \mathbf{U_t}\cdot\mathsf{Diag}\left(\sqrt{\boldsymbol{\lambda}^{(n)}}\right)\cdot(\mathbf{V_P}^{(n)})^H$.

Optimizing variables alternatively, the two-step algorithm converges to the globally optimum solution if the bound is strictly concave on the right singular vectors or the precoder is designed at low SNR region, where the objective function is independent of $\mathbf{V_P}$. When the conditions fail to hold, the iterative algorithm converges to a local maximum, which may be affected by the initialization of the algorithm. However, numerical examples in the next section demonstrate that various initializations have limited effect on the solution; that is, the two-step algorithm achieves near global optimal performance.

V. Simulation Results

Examples are provided in this section to illustrate the relationship and the computational complexity comparison between average mutual information and the derived bounds. They also show the convergence of the proposed algorithm and the efficacy of the designed linear precoder. To exemplify the results, the examples consider the exponential correlation model

$$[\mathbf{\Psi}(\rho)]_{i,j} = \rho^{|i-j|}, \quad \rho \in [0,1) \quad \text{and} \quad i,j = 1,2,\ldots,N$$

with $\mathbf{\Psi_t} = \mathbf{\Psi}(\rho_t)$ and $\mathbf{\Psi_r} = \mathbf{\Psi}(\rho_r)$, as well as the multiple element transmit-receive antennas (METRA) model [24]. The exponential correlation model is suitable for the case of an equally spaced linear array, and the METRA model is verified by measured data in both microcell and picocell environments.

A. Example 1: Relationship Between Average Mutual Information and Bounds

This example is utilized to illustrate that the lower bound (14) plus a constant is a very accurate approximation to the average mutual information.

The illustration begins with the consideration of limits of the average mutual information. When the SNR approaches 0 and $+\infty$ (i.e., σ^2 approaches $+\infty$ and 0, respectively), the limits are given by

$$\lim_{\mathrm{SNR} \longrightarrow 0} \mathsf{E}_{\mathbf{H}} \mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H}) = 0$$
$$\lim_{\mathrm{SNR} \longrightarrow +\infty} \mathsf{E}_{\mathbf{H}} \mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H}) = N_t \log M.$$

At the same time, the limits of the lower bound in (14) are written as

$$\lim_{\mathrm{SNR} \longrightarrow 0} \mathcal{I}_L = -N_r \left(\frac{1}{\ln(2)} - 1 \right)$$
$$\lim_{\mathrm{SNR} \longrightarrow +\infty} \mathcal{I}_L = N_t \log M - N_r \left(\frac{1}{\ln(2)} - 1 \right)$$

which imply a constant gap exists between the average mutual information and the lower bound at low and high SNR regions. Since adding a constant value to the bound remains the optimized precoder unchanged, this work demonstrates that the lower bound with a constant shift actually serves as a very good approximation to the average mutual information. Fig. 1 considers the case with QPSK inputs for exponentially correlated ($\rho_t = 0.8$, $\rho_r = 0.8$) MIMO channels when $N_t = N_r = 2$. The simulated curve is obtained by the Monte Carlo method, which calculates the average mutual information using many realizations of $\mathbf{H_w}$ and $\mathbf{n}$ (see Section V-B for more details). It also shows the derived lower and upper bounds and lower bound with a constant shift for the case of without precoding. With a constant shift, the lower bound and the simulated average mutual information match exactly at low and high SNR regions, and they close to each other at medium SNR region.

Fig. 2 compares further the simulated average mutual information and the lower bound with a constant shift for various numbers of transmit and receive antennas and various input types and correlation parameters. In these cases, the lower bound with a constant shift offers a very good approximation to the average mutual information; the computational complexity for both metric, however, is radically different as shown in the following example.

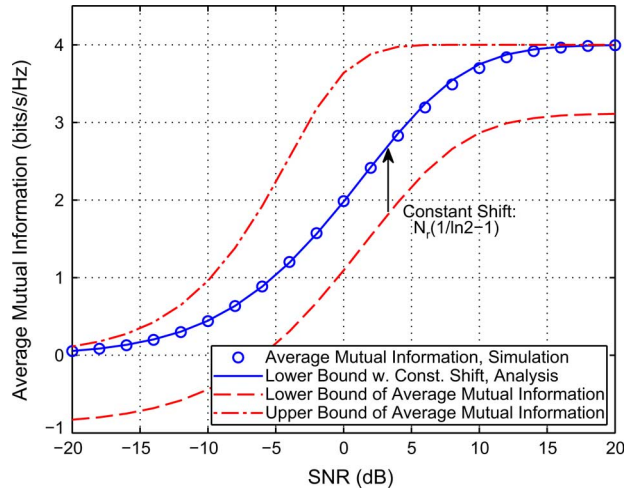


Fig. 1. Average mutual information with QPSK inputs and exponentially correlated ($\rho_t = 0.8$, $\rho_r = 0.8$) MIMO channels ($N_t = N_r = 2$) without precoding.

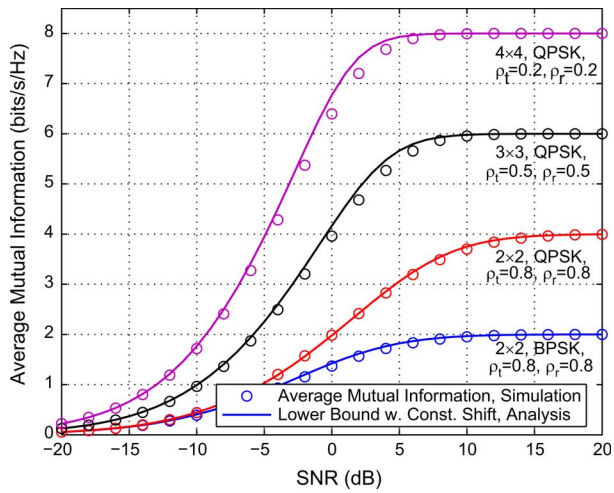


Fig. 2. Average Mutual information with various numbers of transmit and receive antennas and different input types and correlation parameters for the case of without precoding.

B. Example 2: Comparison in Terms of Computational Complexity

This example is employed to demonstrate that the computational complexity of the lower bound is several orders of magnitudes lower than that of the average mutual information.

It is known that the average mutual information lacks closed-form expression and involves multiple integrals. It can be estimated by the Monte Carlo method, which, however, may not be accurate, especially when the number of sample points of noise and channel [see (3) and (4)] are not large enough. Of course, the accuracy can be improved by increasing the number of sample points, which also, unfortunately, increase the computational effort. To make a fair comparison, we choose the number of sample points that cannot further significantly improve the accuracy of the estimated average mutual information. We construct three sequences: the sample number sequence $N = [0, k, 2k, \ldots]$, the average mutual information sequence $A = [A_0, A_1, A_2, \ldots]$ with $A_0 = 0$, and the computation time sequence $T = [T_0, T_1, T_2, \ldots]$ with $T_0 = 0$, where

TABLE III
THE CPU TIME (SECONDS) FOR CALCULATING LOWER BOUND $\mathcal{I}_L$ AND AVERAGE MUTUAL INFORMATION $\mathcal{I}_A$ BASED ON MONTE CARLO METHOD. THE SYMBOL $\times$ INDICATES THE CALCULATION TIME IS MORE THAN ONE HOUR

Cases	BPSK		QPSK	
	Calculate $\mathcal{I}_L$	Calculate $\mathcal{I}_A$	Calculate $\mathcal{I}_L$	Calculate $\mathcal{I}_A$
2×2 MIMO Channels	0.000030	9.369150	0.000049	678.844615
4×4 MIMO Channels	0.000041	406.015399	0.004163	$\times$
6×6 MIMO Channels	0.000365	$\times$	1.705958	$\times$
8×8 MIMO Channels	0.006716	$\times$	450.630528	$\times$
10×10 MIMO Channels	0.155704	$\times$	$\times$	$\times$
12×12 MIMO Channels	3.646632	$\times$	$\times$	$\times$
14×14 MIMO Channels	77.444873	$\times$	$\times$	$\times$
16×16 MIMO Channels	1607.453623	$\times$	$\times$	$\times$

A_i is the estimated average mutual information with T_i seconds using $k \cdot i$ sample points for noise and channel realizations, respectively. The minimal i that makes the improvement of accuracy $|A_i - A_{i-1}|$ less than a threshold α is denoted as $i^\star$:

$$i^\star = \arg\min_i \{|A_i - A_{i-1}| < \alpha\}.$$

Then, the computation time $T_{i^\star}$ is reported.

In our simulations, we set $k = 1000$ and $\alpha = 0.01$. The codes for calculating the bounds and the average mutual information based on Monte Carlo method are written by C++ and are executed on an Intel Core i7–2600 3.40 GHz processor. The typical CPU time with different number of antennas for BPSK and QPSK inputs is shown in Table III, which shows that the computational effort for calculating lower bound is *several orders of magnitude less* than that of calculating the average mutual information. For example, the CPU time of evaluating $\mathcal{I}_L$ for 2×2 MIMO channels with BPSK and 2×2 MIMO channels with QPSK is about 3.2×10^{-6} and 7.2×10^{-8} times that of evaluating $\mathcal{I}_A$, respectively.

Considering the low computational complexity and the accurate approximation, maximizing the lower bound is a reasonable alternative and is expected to offer a very good precoder that maximizes the average mutual information.

C. Example 3: Convergence of the Two-Step Algorithm

This subsection considers the convergence of the two-step method from various feasible initial points and takes the 2×2 exponentially correlated ($\rho_t = 0.95$, $\rho_r = 0.5$) MIMO channels as an example.

A feasible initial point for power allocation vector $\boldsymbol{\lambda}^{(0)}$ is component-wise non-negative and satisfies sum power constraint, whereas a feasible initial point for the right singular vectors $\mathbf{V}_{\mathbf{P}}^{(0)}$ satisfies unitary constraint. A general 2×2 unitary matrix group can be expressed as [25]

$$\mathbf{V} = \underbrace{\begin{bmatrix} e^{\jmath\alpha_1} & 0 \\ 0 & e^{\jmath\alpha_2} \end{bmatrix}}_{\mathbf{V}_1} \cdot \underbrace{\begin{bmatrix} \cos\psi & e^{-\jmath\phi}\sin\psi \\ -e^{\jmath\phi}\sin\psi & \cos\psi \end{bmatrix}}_{\mathbf{V}_2} \tag{50}$$

where $\mathbf{V}_1$ is a diagonal matrix, and $\mathbf{V}_2$ is a unitary matrix with $-\pi < \psi \leq \pi$ and $-\pi/2 \leq \phi \leq \pi/2$. For the simplified channel model (13), the average mutual information remains unchanged under a rotation of $\mathbf{V}_1$. Therefore, the structure of $\mathbf{V}_2$ is used to generate feasible right singular vectors $\mathbf{V}_{\mathbf{P}}^{(0)}$.

The evolution of the proposed two-step algorithm is first considered based on two different initial points: Case A: $\boldsymbol{\lambda} = [1; 1]$, $\psi = \pi/10$, and $\phi = \pi/10$; Case B: $\boldsymbol{\lambda} = [1.6; 0.4]$, $\psi = \pi/6$, and $\phi = \pi/4$. As the lower bound is optimized iteratively by Algorithm 3, the average mutual information is able to increase. Fig. 3 illustrates the corresponding average mutual information of each iteration with both initial points when the input signal is drawn from QPSK, and the SNR is -2.5 dB. For comparison, it also shows the average mutual information without precoding, the maximum capacity with Gaussian inputs for statistical CSI [26], [27], and the precoder optimizing the lower bound directly by gradient descent method. From different initial points, the proposed algorithm converges to almost the same objective value, which increases the average mutual information over the case of without precoding with about 38.7% improvement; thus, it brings the performance of MIMO systems with QPSK inputs close to the maximum capacity with Gaussian inputs. The progress of the proposed algorithm has a staircase shape, with each stair associated with either the iteration within Algorithm 1 for the parameter t or the shift between the optimizations of the power allocation vector and the right singular vectors. The gradient method, however, is influenced by its initial points and converges to different average mutual information levels with different initial points.

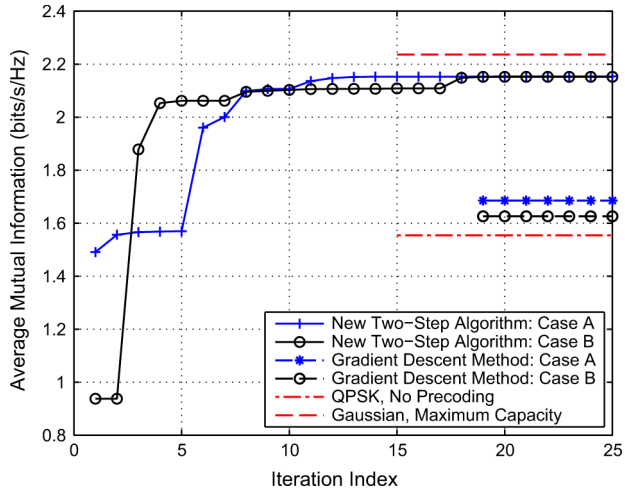


Fig. 3. Typical evolution of average mutual information as the linear precoder is iteratively optimized with the two-step algorithm and gradient descent method. The input signal is drawn from QPSK; 2×2 MIMO channels are exponentially correlated with $\rho_t = 0.95$ and $\rho_r = 0.5$; SNR is -2.5 dB.

The cumulative distribution of average mutual information for the optimized linear precoder from various initial points are further depicted in Fig. 4, which is obtained by generating 10,000 uniform random initial points (applying a normalization

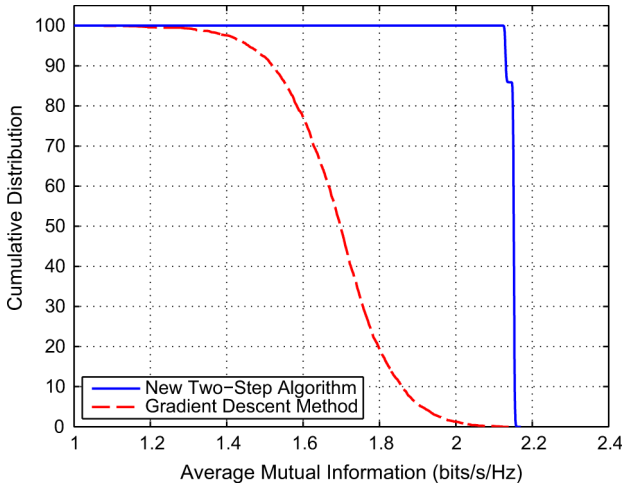


Fig. 4. Cumulative distribution of average mutual information for various initial points. The input signal is drawn from QPSK; 2×2 MIMO channels are exponentially correlated with $\rho_t = 0.95$ and $\rho_r = 0.5$; SNR is -2.5 dB.

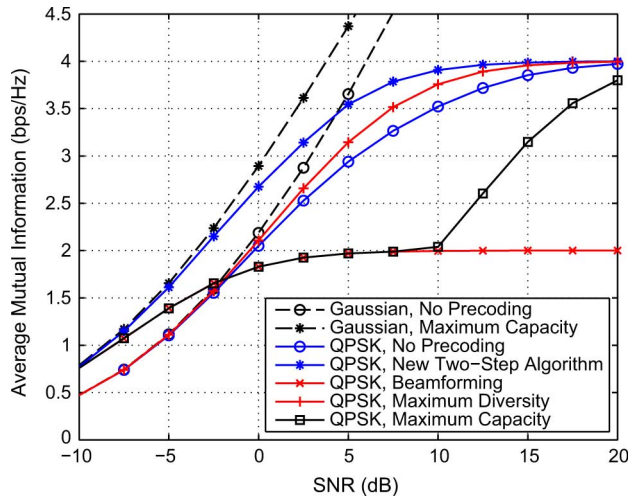


Fig. 5. Average mutual information versus SNR for various strategies. The input signal is drawn from QPSK; 2×2 MIMO channels are exponentially correlated with $\rho_t = 0.95$ and $\rho_r = 0.5$.

for the power allocation vector to guarantee the sum power constraint is satisfied). The cumulative distribution curve illustrates the merit of the proposed algorithm. Although the precoding problem addressed in this work is nonconvex, the two-step algorithm achieves, from an arbitrary initial points, a near global optimal solution, while the gradient descent method depends highly on the selection of the initial points and may provide a solution that performs even worse than the case of no precoding if the initial point is not chosen carefully.

D. Example 4: Efficacy of the Linear Precoder

This example compares the performance of the proposed precoding algorithm with that of other existing algorithms such as beamforming [26], maximum capacity design [26], [27], and maximum diversity [28], [29]. The channel parameters are chosen the same as those in Example 2, and inputs is QPSK modulation. No precoding for both QPSK and Gaussian inputs are also included for comparison purpose.

The average mutual information versus SNR curves of the aforementioned precoding algorithms are depicted in Fig. 5. Obviously, when the channel coding rate is 1/2, the proposed two-step algorithm has about 2.6-, 2.9-, and 11.3-dB gain against the maximum diversity method, no precoding, and the maximum capacity method, respectively. Moreover, when the SNR is less than -2.5 dB, the new algorithm with QPSK input performs almost the same as the maximum capacity method with Gaussian input, which is the ultimate upper bound for all possible linear precoders.

We would like to make two remarks on the maximum capacity precoding method. First, although it obtains performance gain when the channel input signal is Gaussian distribution, it leads to significant performance *loss* if it is directly applied to finite-alphabet signalling. Second, the non-smoothness of the QPSK maximum capacity curve in Fig. 5 (and Fig. 8 to be shown later) is due to the signal mismatch of the precoder designing method. Specifically, when the SNR is less than a threshold (in this case, 7.5 dB), the maximum capacity based precoder allocates all the available power to the strongest subchannels and allocates nothing to the weaker subchannels. This is known as beamforming strategy, which leads to the average mutual information less than 2 bps/Hz. When the SNR is larger than 7.5 dB, the maximum capacity method starts to allocate power to the weaker subchannel and therefore makes the average mutual information larger than 2 b/s/Hz. Since it always allocates more power to the stronger subchannel and less to the weaker subchannel, this method can be far away from the optimal precoder designed for finite-alphabet inputs. The reason is that the average mutual information with finite-alphabet inputs is bounded and therefore it has no incentive to allocate more power to subchannels that already close to saturation. Moreover, the right singular vectors for Gaussian inputs are arbitrary unitary matrices because the statistics of inputs is unchanged when Gaussian signal is rotated; in contrast, the case of finite-alphabet inputs does not follow the same rule, as shown in (6).

We are now in a position to compare the coded bit error rate (BER) performance of the aforementioned precoding algorithms. The transceiver structure [6] in Fig. 6 is then realized to further illustrate the benefit of the proposed linear precoding algorithm. Note that the interleaver is not shown in the block diagram because of the usage of the low-density parity-check, or LDPC, codes [30]. At the transmitter, the signal sequences are encoded by the LDPC encoder and mapped by the conventional equiprobable discrete constellations, respectively. They are then precoded by $\mathbf{P}$ and transmitted through N_t antennas. At the receiver, the maximum *a posteriori* (MAP) detector takes channel observations $\mathbf{y}$ and *a priori* knowledge from the decoder and computes new information for each coded bit. Thus, the extrinsic information between the MAP detector and the decoder is exchanged iteratively until the desired performance is achieved.

The LDPC encoder and decoder modules are derived from [31] with coding rate 1/2. The iteration between the MAP detector and the LDPC decoder is 5. The coding block length of LDPC is 7,200, which includes 600 channel realizations. Multiple blocks are simulated to obtain the coded BER, shown in Fig. 7, of the corresponding precoder schemes in Fig. 5. Apparently, the proposed linear precoding algorithm outperforms all the existing methods, and the BER performance gain over

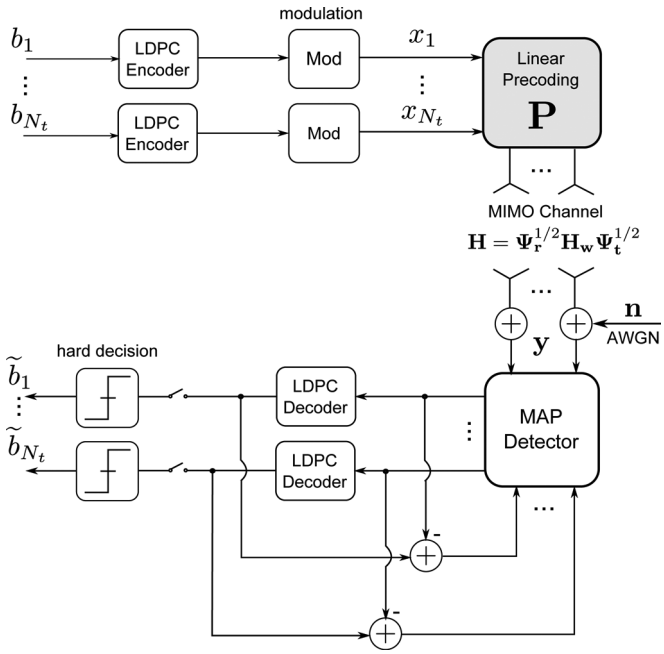


Fig. 6. Block diagram of a MIMO system with channel coding and precoding at the transmitter and iterative detection and decoding at the receiver.

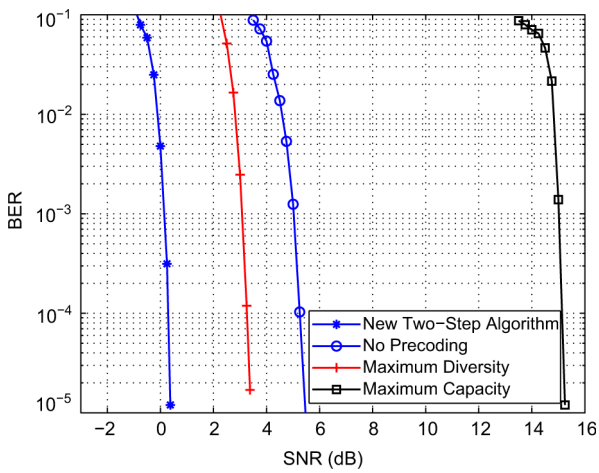


Fig. 7. BER versus SNR for various strategies. The input signal is drawn from QPSK, the channel coding rate is 1/2, coding block length is 7,200 including 600 channel realizations, and the iteration between the MAP detector and the LDPC decoder is 5. MIMO channels of 2×2 are exponentially correlated with $\rho_t = 0.95$ and $\rho_r = 0.5$.

other schemes almost matches the value predicted by the average mutual information in Fig. 5. Based on these results, the large performance gain in average mutual information implies the large performance gain in coded BER. That is, the precoder design based on the average mutual information maximization is an excellent approach to provide considerable performance gain in a practical wireless MIMO system.

The 4×4 MIMO channels with METRA correlation model remains to be considered. Figs. 8 and 9 show the average mutual information based on correlation models for microcell and picocell environments [24], respectively. The microcell model corresponds to a case in which the transmitter is highly correlated, and the picocell model corresponds to a partially decorrelated case (e.g., a small office environment).

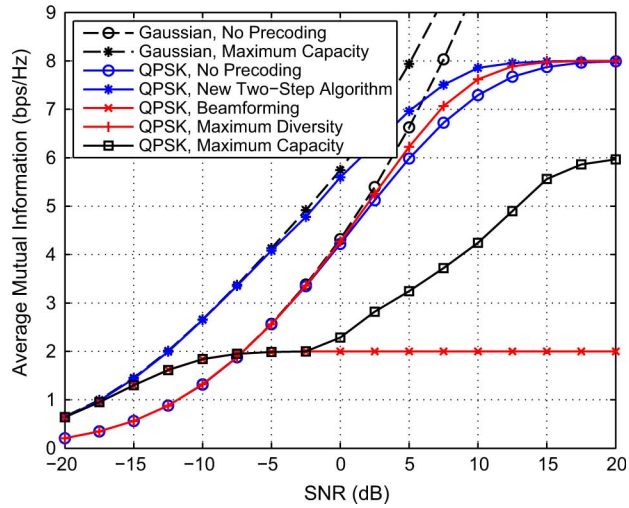


Fig. 8. Average mutual information versus SNR for various strategies. The input signal is drawn from QPSK; 4×4 MIMO channels are generated based on a METRA correlation model for a microcell environment.

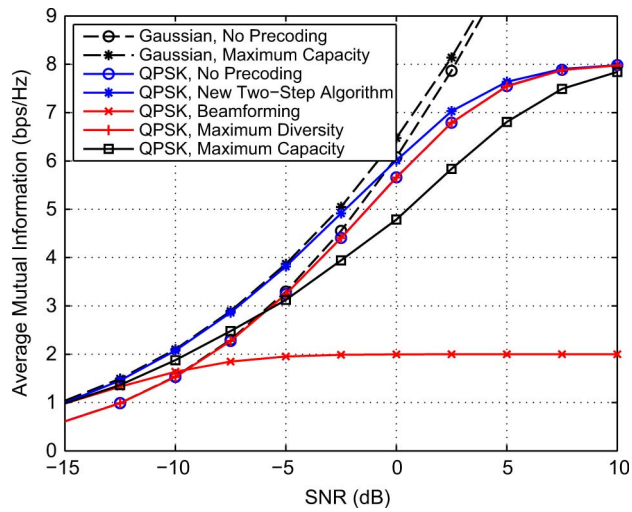


Fig. 9. Average mutual information versus SNR for various strategies. The input signal is drawn from QPSK; 4×4 MIMO channels are generated based on a METRA correlation model for a picocell environment.

Fig. 8 indicates that applying the Gaussian-input based maximum capacity scheme to finite-alphabet inputs results in a significant performance loss for the microcell correlation model when the SNR is larger than -17.5 dB. The proposed algorithm yields a considerable gain in a wide range of SNR; for example, its performance is about 4.3, 4.5, and 14 dB better than the maximum diversity method, no precoding, and the maximum capacity method, respectively, when the channel coding rate is 1/2. At the same time, its performance is the same as that of the maximum capacity method with Gaussian inputs when the SNR is less than -5 dB; that is, it achieves the upper bound for all possible linear precoder designs.

Fig. 9 compares various methods as applied to a picocell correlation environment. Although the proposed algorithm still achieves the upper bound when SNR is less than -5 dB (i.e., providing the same performance as the maximum capacity method with Gaussian inputs), the gain between precoding and no precoding for the picocell case is less than the gain for the microcell model; that is, the maximum gain provided by linear precoder depends on channel environments. Intuitively, when

the channel is statistically independent (i.e., the correlation of transmitter and receiver is identity matrix), the optimal precoder is an arbitrary unitary matrix, which, however, remains the average mutual information unchanged.

VI. Conclusion

In this paper, we have studied the linear precoding method for MIMO wireless fading channels with finite-alphabet inputs, while the transmitter does not know the instantaneous CSI but knows the statistical CSI such as the spatial correlation of the MIMO fading channels. A lower bound and an upper bound have been derived for the average mutual information with finite-alphabet inputs. It is discovered that the lower bound plus a constant is a very accurate approximation to the average mutual information, and the computational complexity of the lower bound is several orders of magnitudes less than that of the average mutual information, which makes the design of optimal linear precoder possible for a practical wireless MIMO system with finite-alphabet inputs. Furthermore, a new two-step iterative algorithm has been proposed to find an optimal linear precoder via capitalizing the derived lower bound.

Numerical examples have demonstrated the convergence and performance gains of the proposed algorithm. Compared to existing linear precoding methods, the proposed linear precoding algorithm offers a significant performance gain in terms of the average mutual information and the coded BER, and the gains become more substantial as the spatial correlation of MIMO channels increases.

Appendix A
Proof of Proposition 1

Before proceeding to the proof, we need to first establish the following lemma.

Lemma 1: The instantaneous mutual information $\mathcal{I}(\mathbf{x};\tilde{\mathbf{y}}|\tilde{\mathbf{H}})$ depends on the precoding matrix $\mathbf{P}$ through $\mathbf{M}=\mathbf{P}^H\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}\mathbf{P}$.

Proof: Similar results are reported in [12]–[14] that assume the transmitter knows perfect CSI. The proof provided here can be seen as an alternative, to the existing proofs, which is based on the reformulation of (4):

$$\begin{aligned}\mathcal{I}(\mathbf{x};\tilde{\mathbf{y}}|\tilde{\mathbf{H}}) &= N_t\log M-\frac{N_r}{\ln 2}-\frac{1}{M^{N_t}}\sum_{m=1}^{M^{N_t}}\\ &\quad \mathsf{E}_{\tilde{\mathbf{n}}}\log\sum_{k=1}^{M^{N_t}}\exp\left(-\frac{\|\tilde{\mathbf{H}}\mathbf{P}\mathbf{e}_{mk}+\tilde{\mathbf{n}}\|^2}{\sigma^2}\right)\\ &= N_t\log M-\frac{N_r}{\ln 2}-\frac{1}{M^{N_t}}\sum_{m=1}^{M^{N_t}}\\ &\quad \mathsf{E}_{l_{mk}}\log\sum_{k=1}^{M^{N_t}}\exp\left(\frac{-l_{mk}}{\sigma^2}\right).\end{aligned}\tag{51}$$

Given $\mathbf{e}_{mk}$ and the SNR, $\mathcal{I}(\mathbf{x};\tilde{\mathbf{y}}|\tilde{\mathbf{H}})$ is a function of the random variable l_{mk}:

$$\begin{aligned}l_{mk}&=\|\tilde{\mathbf{H}}\mathbf{P}\mathbf{e}_{mk}+\tilde{\mathbf{n}}\|^2\\ &=\mathsf{Tr}\left[\mathbf{e}_{mk}\mathbf{e}_{mk}^H\mathbf{P}^H\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}\mathbf{P}+2\Re\left(\mathbf{e}_{mk}^H\mathbf{P}^H\tilde{\mathbf{H}}^H\tilde{\mathbf{n}}\right)+\tilde{\mathbf{n}}\tilde{\mathbf{n}}^H\right],\end{aligned}\tag{52}$$

for $m,k\in\{1,\ldots,M^{N_t}\}$; $\mathcal{I}(\mathbf{x};\tilde{\mathbf{y}}|\tilde{\mathbf{H}})$ changes based on the distribution of l_{mk}, which depends on $\mathbf{P}$ through $\mathbf{P}^H\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}\mathbf{P}$: the first term of the right-hand side in (52) depends on $\mathbf{P}$ through $\mathbf{P}^H\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}\mathbf{P}$; the second term, $\mathbf{e}_{mk}^H\mathbf{P}^H\tilde{\mathbf{H}}^H\mathbf{n}$, is a Gaussian random variable determined by its zero mean and variance $\sigma^2\mathbf{e}_{mk}^H\mathbf{P}^H\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}\mathbf{P}\mathbf{e}_{mk}$ (which also depends on $\mathbf{P}$ through $\mathbf{P}^H\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}\mathbf{P}$); the last term is independent of $\mathbf{P}$. Therefore, $\mathcal{I}(\mathbf{x};\tilde{\mathbf{y}}|\tilde{\mathbf{H}})$ also depends on $\mathbf{P}$ through $\mathbf{M}=\mathbf{P}^H\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}\mathbf{P}$. □

We are now is a position to prove the proposition.

Proof of Proposition 1: Instantaneous mutual information $\mathcal{I}(\mathbf{x};\tilde{\mathbf{y}}|\tilde{\mathbf{H}})$ depends on $\mathbf{P}$ through $\mathbf{M}=\mathbf{P}^H\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}\mathbf{P}$, which is a function of the random matrix $\tilde{\mathbf{H}}_{\mathbf{w}}$. The average mutual information taking expectation over $\tilde{\mathbf{H}}_{\mathbf{w}}$ is thus equal to

$$\mathsf{E}_{\tilde{\mathbf{H}}_{\mathbf{w}}}\hat{\mathcal{I}}(\mathbf{M})=\mathsf{E}_{\mathbf{M}}\hat{\mathcal{I}}(\mathbf{M})\tag{53}$$

where $\hat{\mathcal{I}}(\mathbf{M})$ is used to emphasize the dependence of $\mathcal{I}(\mathbf{x};\tilde{\mathbf{y}}|\tilde{\mathbf{H}})$ on its variable $\mathbf{M}$; that is, the average mutual information is a function of the complex random matrix $\mathbf{M}$ (i.e., $\mathbf{P}^H\mathbf{U}_{\mathbf{t}}\boldsymbol{\Sigma}_{\mathbf{t}}^{\frac{1}{2}}\tilde{\mathbf{H}}_{\mathbf{w}}^H\boldsymbol{\Sigma}_{\mathbf{r}}\tilde{\mathbf{H}}_{\mathbf{w}}\boldsymbol{\Sigma}_{\mathbf{t}}^{\frac{1}{2}}\mathbf{U}_{\mathbf{t}}^H\mathbf{P}$) and thus changes based on the distribution of $\mathbf{M}$.

For the quadratic form $\mathbf{M}$, its probability density function is known as [32]

$$\begin{aligned}p(\mathbf{M})=\frac{1}{\tilde{\Gamma}_{N_r}(N_t)}&\det(\mathbf{W})^{-N_t}\det(\boldsymbol{\Sigma}_{\mathbf{r}})^{-N_r}\det(\mathbf{M})^{N_t-N_r}\\ &\times{}_0\tilde{F}_0^{(N_t)}\left(-\mathbf{W}^{-1}\mathbf{M},\boldsymbol{\Sigma}_{\mathbf{r}}^{-1}\right),\quad \mathbf{M}>0\end{aligned}\tag{54}$$

with

$$\mathbf{W}=\mathbf{P}^H\mathbf{U}_{\mathbf{t}}\boldsymbol{\Sigma}_{\mathbf{t}}\mathbf{U}_{\mathbf{t}}^H\mathbf{P}$$

and

$$\tilde{\Gamma}_{N_r}(\alpha)=\pi^{\frac{N_r(N_r-1)}{2}}\prod_{i=1}^{N_r}\Gamma(\alpha-i+1),\quad \alpha>N_r-1$$

where $\tilde{\Gamma}_{N_r}(\cdot)$ is the complex multivariate gamma function, $\Gamma(\cdot)$ is Euler's gamma function, and ${}_0\tilde{F}_0^{(N_t)}(\cdot)$ is the hypergeometric function of two Hermitian matrices.

Based on (54), the distribution of $\mathbf{M}$ is determined by parameters $\boldsymbol{\Sigma}_{\mathbf{r}}$ and $\mathbf{W}$. Given $\mathbf{W}=\mathbf{P}^H\mathbf{U}_{\mathbf{t}}\boldsymbol{\Sigma}_{\mathbf{t}}\mathbf{U}_{\mathbf{t}}^H\mathbf{P}$, the precoder in the form $\mathbf{P}=\mathbf{U}_{\mathbf{t}}\boldsymbol{\Sigma}_{\mathbf{P}}\mathbf{V}_{\mathbf{P}}^H$ can be used to minimize the transmit power $\mathsf{Tr}(\mathbf{P}\mathbf{P}^H)$; see [33, App. 3.B]. The proof is now complete. □

Appendix B
Proof of Proposition 2

We first prove the lower bound introduced in (14):

Proof of Lower Bound: Note that $\log(x)$ is a concave function in x for $x>0$. Using Jensen's inequality [20], the average mutual information taking expectation on (51) over $\tilde{\mathbf{H}}_{\mathbf{w}}$ can be lower bounded:

$$\begin{aligned}\mathcal{I}_A(\tilde{\mathbf{P}})&\geq N_t\log M-\frac{N_r}{\ln 2}\\ &\quad-\frac{1}{M^{N_t}}\sum_{m=1}^{M^{N_t}}\log\sum_{k=1}^{M^{N_t}}\mathsf{E}_{\tilde{\mathbf{H}}_{\mathbf{w}}}\mathsf{E}_{\mathbf{n}}\exp\left(-\frac{\|\mathbf{c}_{mk}+\mathbf{n}\|^2}{\sigma^2}\right)\end{aligned}\tag{55}$$

where $\mathbf{c}_{mk}=\boldsymbol{\Sigma}_{\mathbf{r}}^{\frac{1}{2}}\tilde{\mathbf{H}}_{\mathbf{w}}\boldsymbol{\Sigma}_{\mathbf{t}}^{\frac{1}{2}}\tilde{\mathbf{P}}\mathbf{e}_{mk}$.

Since $\mathbf{n}$ is an i.i.d Gaussian noise vector, the expectation over $\mathbf{n}$ in (55) can be derived as

$$\begin{aligned}
&\mathsf{E}_{\mathbf{n}} \exp\left(-\frac{\|\mathbf{c}_{mk}+\mathbf{n}\|^2}{\sigma^2}\right) \\
&= \frac{1}{(\pi\sigma^2)^{N_r}} \int_{\mathbf{n}} \exp\left(-\frac{\|\mathbf{c}_{mk}+\mathbf{n}\|^2+\|\mathbf{n}\|^2}{\sigma^2}\right) \mathrm{d}\mathbf{n} \\
&= \frac{1}{(\pi\sigma^2)^{N_r}} \int_{\mathbf{n}} \prod_{i=1}^{N_r} \exp\left(-\frac{\|c_{mk,i}+n_i\|^2+\|n_i\|^2}{\sigma^2}\right) \mathrm{d}\mathbf{n} \\
&= \prod_{i=1}^{N_r} \frac{1}{\pi\sigma^2} \int_{n_i} \exp\left(-\frac{\|c_{mk,i}+n_i\|^2+\|n_i\|^2}{\sigma^2}\right) \mathrm{d}n_i \quad (56)
\end{aligned}$$

where $c_{mk,i}$ and n_i are the ith element of $\mathbf{c}_{mk}$ and $\mathbf{n}$, respectively. Applying the integrals of exponential function [34, eq. (2.33.1)] and extending it to the complex-valued case, (56) is rewritten as

$$\begin{aligned}
\mathsf{E}_{\mathbf{n}} \exp\left(-\frac{\|\mathbf{c}_{mk}+\mathbf{n}\|^2}{\sigma^2}\right) &= \prod_{i=1}^{N_r} \frac{1}{2} \exp\left(-\frac{\|c_{mk,i}\|^2}{2\sigma^2}\right) \\
&= \frac{1}{2^{N_r}} \exp\left(-\frac{\mathbf{c}_{mk}^H \mathbf{c}_{mk}}{2\sigma^2}\right). \quad (57)
\end{aligned}$$

Considering the property $\mathsf{Tr}(\mathbf{AB}) = \mathsf{Tr}(\mathbf{BA})$ and recalling the definition $\mathbf{c}_{mk} = \mathbf{\Sigma}_{\mathbf{r}}^{\frac{1}{2}} \tilde{\mathbf{H}}_{\mathbf{w}} \mathbf{\Sigma}_{\mathbf{t}}^{\frac{1}{2}} \tilde{\mathbf{P}} \mathbf{e}_{mk}$, it yields

$$\exp\left(-\frac{\mathbf{c}_{mk}^H \mathbf{c}_{mk}}{2\sigma^2}\right) = \exp\left(-\frac{\mathsf{Tr}\left(\mathbf{\Sigma}_{\mathbf{r}} \tilde{\mathbf{H}}_{\mathbf{w}} \mathbf{Q}_{mk} \tilde{\mathbf{H}}_{\mathbf{w}}^H\right)}{2\sigma^2}\right) \quad (58)$$

where $\mathbf{Q}_{mk} = \mathbf{\Sigma}_{\mathbf{t}}^{\frac{1}{2}} \tilde{\mathbf{P}} \mathbf{e}_{mk} \mathbf{e}_{mk}^H \tilde{\mathbf{P}}^H \mathbf{\Sigma}_{\mathbf{t}}^{\frac{1}{2}}$. Now denoting the distribution of $\tilde{\mathbf{H}}_{\mathbf{w}}$ by $p(\tilde{\mathbf{H}}_{\mathbf{w}})$, the right-hand side of (55) suggests to compute

$$\begin{aligned}
&\mathsf{E}_{\tilde{\mathbf{H}}_{\mathbf{w}}} \exp\left(-\frac{\mathsf{Tr}\left(\mathbf{\Sigma}_{\mathbf{r}} \tilde{\mathbf{H}}_{\mathbf{w}} \mathbf{Q}_{mk} \tilde{\mathbf{H}}_{\mathbf{w}}^H\right)}{2\sigma^2}\right) \\
&= \int_{\tilde{\mathbf{H}}_{\mathbf{w}}} \exp\left(-\frac{\mathsf{Tr}\left(\mathbf{\Sigma}_{\mathbf{r}} \tilde{\mathbf{H}}_{\mathbf{w}} \mathbf{Q}_{mk} \tilde{\mathbf{H}}_{\mathbf{w}}^H\right)}{2\sigma^2}\right) p(\tilde{\mathbf{H}}_{\mathbf{w}}) \mathrm{d}\tilde{\mathbf{H}}_{\mathbf{w}}. \quad (59)
\end{aligned}$$

To facilitate the integration, $\mathsf{Tr}\left(\mathbf{\Sigma}_{\mathbf{r}} \tilde{\mathbf{H}}_{\mathbf{w}} \mathbf{Q}_{mk} \tilde{\mathbf{H}}_{\mathbf{w}}^H\right)$ is reorganized as a quadratic form in the entries of $\tilde{\mathbf{H}}_{\mathbf{w}}$:

$$\begin{aligned}
&\mathsf{Tr}\left(\mathbf{\Sigma}_{\mathbf{r}} \tilde{\mathbf{H}}_{\mathbf{w}} \mathbf{Q}_{mk} \tilde{\mathbf{H}}_{\mathbf{w}}^H\right) \\
&= \sum_q \sum_i \sum_j r_q \cdot [\tilde{\mathbf{H}}_{\mathbf{w}}]_{q,i} \cdot [\mathbf{Q}_{mk}]_{i,j} \cdot [\tilde{\mathbf{H}}_{\mathbf{w}}]_{q,j}^H \\
&= \sum_q [\tilde{\mathbf{H}}_{\mathbf{w}}]_{q,:} \cdot (r_q \mathbf{Q}_{mk}) \cdot [\tilde{\mathbf{H}}_{\mathbf{w}}]_{q,:}^H \quad (60)
\end{aligned}$$

where r_q denotes the qth diagonal element of $\mathbf{\Sigma}_{\mathbf{r}}$. Letting $\mathbf{h}_q = [\tilde{\mathbf{H}}_{\mathbf{w}}]_{q,:}^H$, the integral over $\tilde{\mathbf{H}}_{\mathbf{w}}$ in (59) is written as the product of several elementary integrals:

$$\begin{aligned}
&\int_{\tilde{\mathbf{H}}_{\mathbf{w}}} \exp\left(-\frac{\mathsf{Tr}\left(\mathbf{\Sigma}_{\mathbf{r}} \tilde{\mathbf{H}}_{\mathbf{w}} \mathbf{Q}_{mk} \tilde{\mathbf{H}}_{\mathbf{w}}^H\right)}{2\sigma^2}\right) p(\tilde{\mathbf{H}}_{\mathbf{w}}) \mathrm{d}\tilde{\mathbf{H}}_{\mathbf{w}} \\
&= \int_{\tilde{\mathbf{H}}_{\mathbf{w}}} \prod_q \frac{1}{\pi^{N_r}} \exp\left(-\mathbf{h}_q^H \left(\mathbf{I} + \frac{r_q}{2\sigma^2} \mathbf{Q}_{mk}\right) \mathbf{h}_q\right) \mathrm{d}\tilde{\mathbf{H}}_{\mathbf{w}} \\
&= \prod_q \left[\frac{1}{\pi^{N_r}} \int_{\mathbf{h}_q} \exp\left(-\mathbf{h}_q^H \left(\mathbf{I} + \frac{r_q}{2\sigma^2} \mathbf{Q}_{mk}\right) \mathbf{h}_q\right) \mathrm{d}\mathbf{h}_q\right]. \quad (61)
\end{aligned}$$

According to the exponential quadratic integral formula [35, eq. (C1)]

$$\int_{\mathbf{h}_q} e^{-\mathbf{h}_q^H (\mathbf{A}+j\mathbf{B}) \mathbf{h}_q} \mathrm{d}\mathbf{h}_q = \frac{\pi^{N_r}}{\det(\mathbf{A}+j\mathbf{B})} \quad (62)$$

with both $\mathbf{A}$ and $\mathbf{B}$ from Hermitian matrices, the expression in (61) is further reduced to

$$\begin{aligned}
&\int_{\tilde{\mathbf{H}}_{\mathbf{w}}} \exp\left(-\frac{\mathsf{Tr}\left(\mathbf{\Sigma}_{\mathbf{r}} \tilde{\mathbf{H}}_{\mathbf{w}} \mathbf{Q}_{mk} \tilde{\mathbf{H}}_{\mathbf{w}}^H\right)}{2\sigma^2}\right) p(\tilde{\mathbf{H}}_{\mathbf{w}}) \mathrm{d}\tilde{\mathbf{H}}_{\mathbf{w}} \\
&\qquad = \prod_q \left[\det\left(\mathbf{I} + \frac{r_q}{2\sigma^2} \mathbf{Q}_{mk}\right)\right]^{-1}. \quad (63)
\end{aligned}$$

Recalling the definition of $\mathbf{Q}_{mk}$, the determinant in (63) is given by

$$\det\left(\mathbf{I} + \frac{r_q}{2\sigma^2} \mathbf{Q}_{mk}\right) = 1 + \frac{r_q}{2\sigma^2} \mathbf{e}_{mk}^H \tilde{\mathbf{P}}^H \mathbf{\Sigma}_{\mathbf{t}} \tilde{\mathbf{P}} \mathbf{e}_{mk}. \quad (64)$$

The combination of (59), (63), and (64) yields the lower bound. □

We then prove the upper bound in (15):

Proof of Upper Bound: Because $\log \sum_k \exp(x_k)$ is a convex function, Jensen's inequality can be applied to the expression of average mutual information to derive an upper bound:

$$\begin{aligned}
\mathcal{I}_A(\tilde{\mathbf{P}}) \le N_t \log M - \frac{1}{M^{N_t}} \sum_{m=1}^{M^{N_t}} \\
\log \sum_{k=1}^{M^{N_t}} \exp\left(-\mathsf{E}_{\tilde{\mathbf{H}}_{\mathbf{w}}} \mathsf{E}_{\mathbf{n}} \frac{\|\mathbf{c}_{mk}+\mathbf{n}\|^2 - \|\mathbf{n}\|^2}{\sigma^2}\right). \quad (65)
\end{aligned}$$

The expectation over $\mathbf{n}$ is first considered:

$$\begin{aligned}
\mathsf{E}_{\mathbf{n}} \frac{\|\mathbf{c}_{mk}+\mathbf{n}\|^2 - \|\mathbf{n}\|^2}{\sigma^2} &= \frac{\mathbf{c}_{mk}^H \mathbf{c}_{mk}}{\sigma^2} \\
&= \frac{\mathbf{e}_{mk}^H \tilde{\mathbf{P}}^H \mathbf{\Sigma}_{\mathbf{t}}^{\frac{1}{2}} \tilde{\mathbf{H}}_{\mathbf{w}}^H \mathbf{\Sigma}_{\mathbf{r}} \tilde{\mathbf{H}}_{\mathbf{w}} \mathbf{\Sigma}_{\mathbf{t}}^{\frac{1}{2}} \tilde{\mathbf{P}} \mathbf{e}_{mk}}{\sigma^2}. \quad (66)
\end{aligned}$$

Then the expectation over $\tilde{\mathbf{H}}_{\mathbf{w}}$ is given by

$$\begin{aligned}
&\mathsf{E}_{\tilde{\mathbf{H}}_{\mathbf{w}}}\mathsf{E}_{\mathbf{n}}\frac{\|\mathbf{c}_{mk}+\mathbf{n}\|^2-\|\mathbf{n}\|^2}{\sigma^2}\\
&\quad=\frac{\mathbf{e}_{mk}^H\tilde{\mathbf{P}}^H\boldsymbol{\Sigma}_{\mathbf{t}}^{\frac{1}{2}}\left(\mathsf{E}_{\tilde{\mathbf{H}}_{\mathbf{w}}}\tilde{\mathbf{H}}_{\mathbf{w}}^H\boldsymbol{\Sigma}_{\mathbf{r}}\tilde{\mathbf{H}}_{\mathbf{w}}\right)\boldsymbol{\Sigma}_{\mathbf{t}}^{\frac{1}{2}}\tilde{\mathbf{P}}\mathbf{e}_{mk}}{\sigma^2}\\
&\quad=\frac{\mathbf{e}_{mk}^H\tilde{\mathbf{P}}^H\boldsymbol{\Sigma}_{\mathbf{t}}^{\frac{1}{2}}\mathsf{Tr}(\boldsymbol{\Sigma}_{\mathbf{r}})\boldsymbol{\Sigma}_{\mathbf{t}}^{\frac{1}{2}}\tilde{\mathbf{P}}\mathbf{e}_{mk}}{\sigma^2}\\
&\quad=\frac{\mathsf{Tr}(\boldsymbol{\Sigma}_{\mathbf{r}})}{\sigma^2}\mathbf{e}_{mk}^H\tilde{\mathbf{P}}^H\boldsymbol{\Sigma}_{\mathbf{t}}\tilde{\mathbf{P}}\mathbf{e}_{mk} \qquad (67)
\end{aligned}$$

which completes the proof. □

APPENDIX C
PROOF OF CONCAVITY RESULTS

Proof of Proposition 5: The concavity of the lower bound function can be proved by reformulating (14) as

$$\begin{aligned}
\mathcal{I}_L(\tilde{\mathbf{P}}) &= N_t\log M - N_r\left(\frac{1}{\ln 2}-1\right)\\
&\quad-\frac{1}{M^{N_t}}\sum_{m=1}^{M^{N_t}}\log\sum_{k=1}^{M^{N_t}}\\
&\quad\exp\left[-\sum_q\ln\left(1+\frac{r_q}{2\sigma^2}\mathsf{Tr}\left(\tilde{\mathbf{P}}^H\boldsymbol{\Sigma}_{\mathbf{t}}\tilde{\mathbf{P}}\mathbf{e}_{mk}\mathbf{e}_{mk}^H\right)\right)\right]. \qquad (68)
\end{aligned}$$

Since $-\log\sum_i\exp(-g_i)$ is concave whenever g_i are concave [20], the concavity of (68) depends on that of $\ln\left(1+\frac{r_q}{2\sigma^2}\mathsf{Tr}(\tilde{\mathbf{P}}^H\boldsymbol{\Sigma}_{\mathbf{t}}\tilde{\mathbf{P}}\mathbf{e}_{mk}\mathbf{e}_{mk}^H)\right)$, which can be proved based on the fact that $\mathsf{Tr}\left(\tilde{\mathbf{P}}^H\boldsymbol{\Sigma}_{\mathbf{t}}\tilde{\mathbf{P}}\mathbf{e}_{mk}\mathbf{e}_{mk}^H\right)$ is positive and concave over $\tilde{\mathbf{P}}^H\boldsymbol{\Sigma}_{\mathbf{t}}\tilde{\mathbf{P}}$. Similar steps can be followed to prove upper bound function (15).

Proof of Proposition 6: Following a method similar to that used to derive Proposition 5, the concavity here can be proved by re-expressing the lower bound of average mutual information as

$$\begin{aligned}
\mathcal{I}_L(\tilde{\mathbf{P}}) &= N_t\log M - N_r\left(\frac{1}{\ln 2}-1\right)-\frac{1}{M^{N_t}}\sum_{m=1}^{M^{N_t}}\log\sum_{k=1}^{M^{N_t}}\\
&\quad\exp\left[-\sum_q\ln\left(1+\frac{r_q}{2\sigma^2}\right.\right.\\
&\quad\left.\left.\times\,\mathsf{Tr}\left(\boldsymbol{\Sigma}_{\mathbf{t}}\mathbf{V}_{\mathbf{P}}^H\mathbf{e}_{mk}\mathbf{e}_{mk}^H\mathbf{V}_{\mathbf{P}}\boldsymbol{\Sigma}_{\mathbf{P}}^2\right)\right)\right] \qquad (69)
\end{aligned}$$

and confirming that $\mathsf{Tr}\left(\boldsymbol{\Sigma}_{\mathbf{t}}\mathbf{V}_{\mathbf{P}}^H\mathbf{e}_{mk}\mathbf{e}_{mk}^H\mathbf{V}_{\mathbf{P}}\boldsymbol{\Sigma}_{\mathbf{P}}^2\right)$ is positive and concave over $\boldsymbol{\Sigma}_{\mathbf{P}}^2$.

Since the power allocation vector $\boldsymbol{\lambda}$ is given by

$$\boldsymbol{\lambda}=\mathsf{diag}(\boldsymbol{\Sigma}_{\mathbf{P}}^2) \qquad (70)$$

the gradient with respect to $\boldsymbol{\lambda}$ can be written as

$$\nabla_{\boldsymbol{\lambda}}\mathcal{I}_L=\mathsf{diag}\left(\frac{\partial\mathcal{I}_L}{\partial\boldsymbol{\Sigma}_{\mathbf{P}}^2}\right). \qquad (71)$$

Applying the techniques for matrix differentiation [36], the gradient of (69) with respect to $\boldsymbol{\lambda}$ is derived in (28). Using the same method, the gradient of (15) is readily given by (31). □

ACKNOWLEDGMENT

The authors would like to thank the associate editor, Prof. D. Love, and anonymous reviewers for helpful comments and suggestions that greatly improved the quality of the paper.

REFERENCES

[1] A. Scaglione, P. Stoica, S. Barbarossa, G. Giannakis, and H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1051–1064, May 2002.
[2] W. Su and X. Xia, "Signal constellations for quasi-orthogonal space-time block codes with full diversity," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2331–2347, Oct. 2004.
[3] D. Love and R. Heath, "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2967–2976, Aug. 2005.
[4] Y. Rong, S. Vorobyov, and A. Gershman, "Linear block precoding for OFDM systems based on maximization of mean cutoff rate," *IEEE Trans. Signal Process.*, vol. 53, no. 12, pp. 4691–4696, Dec. 2005.
[5] M. Vu and A. Paulraj, "MIMO wireless linear precoding," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 86–105, Sep. 2007.
[6] B. Hochwald and S. TenBrink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, Mar. 2003.
[7] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.
[8] A. Lozano, A. 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.
[9] D. Palomar and S. Verdú, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 141–154, Jan. 2006.
[10] F. Pérez-Cruz, M. 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.
[11] C. Xiao and Y. R. Zheng, "On the mutual information and power allocation for vector Gaussian channels with finite discrete inputs," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, New Orleans, LA, 2008.
[12] 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)*, Seoul, Korea, 2009, pp. 1085–1089.
[13] M. Lamarca, "Linear precoding for mutual information maximization in MIMO systems," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, 2009, pp. 26–30.
[14] C. Xiao, Y. R. Zheng, and Z. Ding, "Globally optimal linear precoders for finite alphabet signals over complex vector Gaussian channels," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3301–3314, Jul. 2011.
[15] O. Oyman, R. Nabar, H. Bolcskei, and A. Paulraj, "Characterizing the statistical properties of mutual information in MIMO channels," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2784–2795, Nov. 2003.
[16] E. Jorswieck and H. Boche, "Optimal transmission strategies and impact of correlation in multiantenna systems with different types of channel state information," *IEEE Trans. Signal Process.*, vol. 52, no. 12, pp. 3440–3453, Dec. 2004.
[17] E. Yoon, J. Hansen, and A. Paulraj, "Space-frequency precoding with space-tap correlation information at the transmitter," *IEEE Trans. Commun.*, vol. 55, no. 9, pp. 1702–1711, Sep. 2007.
[18] P. Xiao and M. Sellathurai, "Improved linear transmit processing for single-user and multi-user MIMO communications systems," *IEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1768–1779, Mar. 2010.
[19] C. Xiao, J. Wu, S. Leong, Y. R. Zheng, and K. Letaief, "A discrete-time model for triply selective MIMO Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1678–1688, Sep. 2004.
[20] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[21] M. Payaró and D. P. Palomar, "Hessian and concavity of mutual information, differential entropy, and entropy power in linear vector Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3613–3628, Aug. 2009.
[22] J. H. Manton, "Optimization algorithms exploiting unitary constraints," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 635–650, Mar. 2002.
[23] R. Horn and C. Johnson, *Matrix Analysis Minus*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[24] J. Kermoal, L. Schumacher, K. Pedersen, P. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 6, pp. 1211–1226, Jun. 2002.
[25] F. Murnaghan, *Lectures on Applied Mathematics: The Unitary and Rotation Groups*. Washington, DC: Spartan Books, 1962, vol. 3.
[26] E. Jorswieck and H. Boche, "Cannel capacity and capacity-range of beamforming in MIMO wireless systems under correlated fading with covariance feedback," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1543–1553, May 2004.
[27] A. Soysal and S. Ulukus, "Optimum power allocation for single-user MIMO and multi-user MIMO-MAC with partial CSI," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1402–1412, Jul. 2007.
[28] J. Boutros and E. Viterbo, "Signal space diversity: A power-and bandwidth-efficient diversity technique for the Rayleigh fading channel," *IEEE Trans. Inf. Theory*, vol. 44, no. 4, pp. 1453–1467, Apr. 1998.
[29] Y. Xin, Z. Wang, and G. Giannakis, "Space-time diversity systems based on linear constellation precoding," *IEEE Trans. Wireless Commun.*, vol. 2, no. 2, pp. 294–309, Feb. 2003.
[30] B. Lu, X. Wang, and K. Narayanan, "LDPC-based space-time coded OFDM systems over correlated fading channels: Performance analysis and receiver design," *IEEE Trans. Commun.*, vol. 50, no. 1, pp. 74–88, Jan. 2002.
[31] M. Valenti, Iterative Solutions Coded Modulation Library (ISCML) [Online]. Available: http://www.iterativesolutions.com/Matlab.htm
[32] H. Shin, M. Win, J. Lee, and M. Chiani, "On the capacity of doubly correlated MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 5, no. 8, pp. 2253–2265, Aug. 2006.
[33] D. Palomar and Y. Jiang, *MIMO Transceiver Design via Majorization Theory*. Delft, The Netherlands: Now Publishers, 2006.
[34] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. New York: Academic, 2007.
[35] B. Hassibi and T. Marzetta, "Multiple-antennas and isotropically random unitary inputs: The received signal density in closed form," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1473–1484, Jun. 2002.
[36] J. Magnus and H. Neudecker, *Matrix Differential Calculus With Applications in Statistics and Econometrics*, 3rd ed. New York: Wiley, 2007.

Weiliang Zeng (S'08) received the B.S. degree in electronic engineering (with the Highest Hons.) from the University of Electronic Sciences and Technology of China (UESTC), Chengdu, China, in 2007.

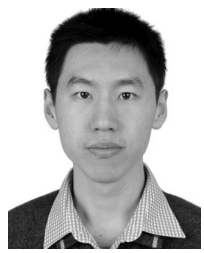
Since September 2007, he has been working trowards the Ph.D. degree at the Tsinghua National Laboratory for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing, China. During his doctoral studies, he conducted cooperative research at the Department of Electronic Engineering, Missouri University of Science and Technology, Rolla. His research interests include communications and information theory with special emphasis on wireless communications and signal processing.

Mr. Zeng is the recipient of first-class scholarship for graduate students at Tsinghua University.

Chengshan Xiao (M'99–SM'02–F'10) received the B.S. degree in electronic engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 1987, the M.S. degree in electronic engineering from Tsinghua University, Beijing, China, in 1989, and the Ph.D. degree in electrical engineering from the University of Sydney, Sydney, Australia, in 1997.

From 1989 to 1993, he was with the Department of Electronic Engineering, Tsinghua University, where he was a member of Research Staff and then a Lecturer. From 1997 to 1999, he was a Senior Member of Scientific Staff with Nortel, Ottawa, ON, Canada. From 1999 to 2000, he was a Faculty Member with the University of Alberta, Edmonton, AB, Canada. From 2000 to 2007, he was with the University of Missouri, Columbia, where he was an Assistant Professor and then an Associate Professor. He is currently a Professor with the Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla (formerly, the University of Missouri). His research interests include wireless communications, signal processing, and underwater acoustic communications. He holds three U.S. patents. His algorithms have been implemented into Nortel's base station radios after successful technical field trials and network integration.

Dr. Xiao is the Editor-in-Chief of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, a Member of the Fellow Evaluation Committee of IEEE Communications Society (ComSoc), a Member at Large of the IEEE ComSoc Board of Governors, a Distinguished Lecturer of IEEE ComSoc, and a Distinguished Lecturer of the IEEE Vehicular Technology Society. Previously, he served as the founding Area Editor for Transmission Technology of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS; an Associate Editor of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I, and the international journal *Multidimensional Systems and Signal Processing*. He was the Technical Program Chair of the 2010 IEEE International Conference on Communications (ICC), the Lead Co-Chair of the 2008 IEEE ICC Wireless Communications Symposium, and a Phy/MAC Program Co-Chair of the 2007 IEEE Wireless Communications and Networking Conference. He served as the founding Chair of the IEEE Technical Committee on Wireless Communications and the Vice-Chair of the IEEE Technical Committee on Personal Communications.



Mingxi Wang (S'10) received the B.S degree in telecommunication engineering and the M.S. degree in signal and information processing from Beihang University (formerly, the Beijing University of Aeronautics and Astronautics), China, in 2006 and 2009, respectively.

He is currently working towards the Ph.D. degree of the Department of Electrical and Computer Engineering at the Missouri University of Science and Technology (formerly, the University of Missouri—Rolla). His research interests include wireless communications, signal processing, MIMO, OFDM, linear precoding, and iterative receiver.

Jianhua Lu (M'98–SM'07) received the B.S.E.E. and M.S.E.E. degrees from Tsinghua University, Beijing, China, in 1986 and 1989, respectively, and the Ph.D. degree in electrical and electronic engineering from the Hong Kong University of Science and Technology, Kowloon.

Since 1989, he has been with the Department of Electronic Engineering, Tsinghua University, where he is currently a Professor. His research interests include broadband wireless communication, multimedia signal processing, satellite communication, and wireless networking. He has published more than 180 technical papers in international journals and conference proceedings.

Dr. Lu has been an active member of professional societies. He was one of the recipients of Best Paper Awards at the International Conference on Communications, Circuits and Systems 2002 and ChinaCom 2006 and was awarded the National Distinguished Young Scholar Fund by the NSF committee of China in 2006. He has served in numerous IEEE conferences as a member of Technical Program Committees and served as Lead Chair of the General Symposium of IEEE ICC 2008, as well as a Program Committee Co-Chair of the Ninth IEEE International Conference on Cognitive Informatics.