Linear Precoding for Fading Cognitive Multiple-Access Wiretap Channel With Finite-Alphabet Inputs

Juening Jin, Chengshan Xiao, *Fellow, IEEE*, Meixia Tao, *Senior Member, IEEE*, and Wen Chen, *Senior Member, IEEE*

***Abstract*—We investigate the fading cognitive multiple-access wiretap channel (CMAC-WT), in which two secondary-user transmitters (STs) send secure messages to a secondary-user receiver (SR) in the presence of an eavesdropper and subject to interference threshold constraints at multiple primary-user receivers (PRs). We design linear precoders to maximize the average secrecy sum rate for a multiple-input–multiple-output (MIMO) fading CMAC-WT under finite-alphabet inputs and statistical channel state information at STs. For this nondeterministic polynomial-time NP-hard problem, we utilize an accurate approximation of the average secrecy sum rate to reduce the computational complexity and then present a two-layer algorithm by embedding the convex–concave procedure into an outer-approximation framework. The idea behind this algorithm is to reformulate the approximated average secrecy sum rate as a difference of convex functions and then generate a sequence of simpler relaxed sets to approach the nonconvex feasible set. Subsequently, we maximize the approximated average secrecy sum rate over the sequence of relaxed sets by using the convex–concave procedure. Numerical results indicate that our proposed precoding algorithm is superior to the conventional Gaussian precoding method in the medium and high signal-to-noise ratio (SNR) regimes.**

***Index Terms*—Cognitive multiple-access wiretap channel (CMAC-WT), finite-alphabet inputs, linear precoding, multiple-input multiple-output (MIMO), physical-layer security, statistical channel state information (CSI).**

I. Introduction

Spectrum sharing has been widely recognized as a promising technology to improve the utilization efficiency of the limited spectrum resources in cognitive radio networks [1]. In a spectrum sharing cognitive radio network, unlicensed secondary users are allowed to concurrently communicate with licensed primary users over the same bandwidth as long as the interference power at primary-user receivers (PRs) is kept below a given threshold. Related works in [2] and [3] considered the weighted sum-rate optimization in cognitive radio networks with interference threshold constraints.

Meanwhile, due to the open and broadcast nature of radio propagation, such spectrum sharing may cause security problems because all kinds of wireless equipment devices are able to overhear the licensed spectrum. Therefore, security is a critical issue in cognitive radio networks. Traditionally, the security of a network has been entrusted in the network layer through cryptography and authentication, which often require additional system complexity for key generation and complex encryption/decryption algorithms [4].

In recent years, there has been growing interest in physical-layer security that enables secure communication over the physical layer. Physical-layer security or information-theoretic security originated from Shannon's notion of perfect secrecy [5]. It was first studied in a wiretap channel by Wyner [6] and later in a broadcast channel with confidential messages by Csiszár and Körner [7]. The study of physical-layer security is then extended to several multiuser communication scenarios. In [8], Tekin and Yener introduced the degraded Gaussian multiple-access wiretap channel, where an additional eavesdropper (ED) is able to access the multiple-access channel output via a degraded wiretap channel. In [9], an achievable secrecy rate region with Gaussian inputs was proposed for the nondegraded Gaussian multiple-access wiretap channel, and the power allocations maximizing the corresponding secrecy sum rate were also determined. Related works in [10]–[14] further investigated linear precoding designs that maximize the secrecy (sum) rate in other multiple-input multiple-output (MIMO) multiuser channels.

The precoding designs in [10]–[14] require instantaneous channel state information (CSI) of both legitimate receivers and EDs. However, such a requirement is overoptimistic for fast-fading channels, of which the channel coherence time may be shorter than the feedback delay caused by channel

Manuscript received November 29, 2015; revised April 26, 2016; accepted June 20, 2016. Date of publication July 12, 2016; date of current version April 14, 2017. The work of C. Xiao was supported by the National Science Foundation under Grant ECCS-1231848 and Grant ECCS-1539316. The work of M. Tao was supported by the National Natural Science Foundation of China under Grant 61322102. The work of W. Chen was supported in part by the National 973 Project under Grant 2012CB316106 and in part by the National 863 Project under Grant 2015AA01A710. This work has been carried out while J. Jin was visiting Missouri University of Science and Technology. This paper was presented in part at the IEEE International Conference on Communications, Kuala Lumpur, Malaysia, May 23–27, 2016. The review of this paper was coordinated by Dr. H. Lin.

J. Jin and M. Tao are with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: jueningjin@gmail.com; mxtao@sjtu.edu.cn).

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

W. Chen is with the Shanghai Key Laboratory of Navigation and Location Based Services, Shanghai Jiao Tong University, Shanghai 200240, China, and also with the School of Electronics Engineering and Automation, Guilin University of Electronics Technology, Guilin 541004, China (e-mail: wenchen@sjtu.edu.cn).

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/TVT.2016.2590539

0018-9545 © 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

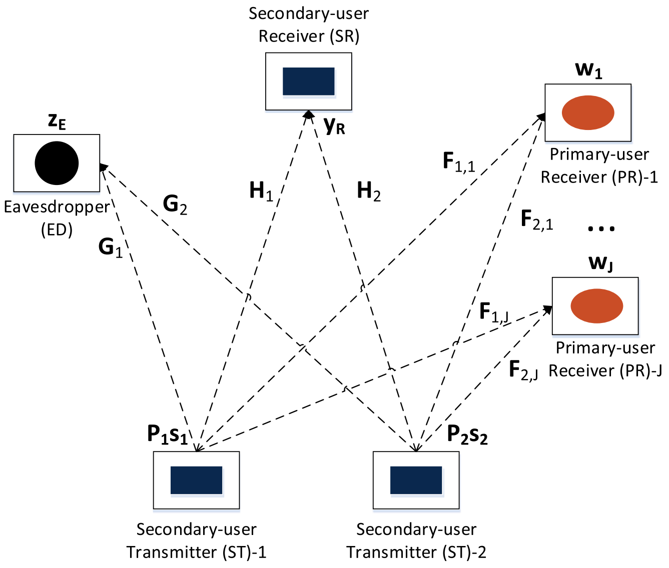

Secondary-user Receiver (SR)
$\mathbf{y}_R$
$\mathbf{z}_E$
Eavesdropper (ED)
$\mathbf{w}_1$
Primary-user Receiver (PR)-1
$\mathbf{w}_J$
Primary-user Receiver (PR)-J
$\mathbf{G}_1$
$\mathbf{G}_2$
$\mathbf{H}_1$
$\mathbf{H}_2$
$\mathbf{F}_{1,1}$
$\mathbf{F}_{2,1}$
$\mathbf{F}_{1,J}$
$\mathbf{F}_{2,J}$
…
$\mathbf{P}_1\mathbf{s}_1$
$\mathbf{P}_2\mathbf{s}_2$
Secondary-user Transmitter (ST)-1
Secondary-user Transmitter (ST)-2

Fig. 1. System model of the fading CMAC-WT.

estimation. In this case, when the instantaneous CSI arrives at transmitters, the channel state has already changed. Therefore, it is more realistic to exploit channel statistics at transmitters for precoding design, due to its much slower changes compared with instantaneous CSI.

Furthermore, the results in [10]–[14] rely on the ideal assumption of Gaussian inputs. Although Gaussian inputs are proven to be capacity achieving in a variety of Gaussian channels, they are hardly implemented in practice. It is well known that practical inputs are drawn from finite constellation sets such as phase-shift keying (PSK), pulse-amplitude modulation, or quadrature amplitude modulation. More importantly, the common approach of designing a linear precoder in a MIMO system under Gaussian inputs and then applying it to the practical system may lead to significant performance loss [15], [16]. Therefore, the precoding design with finite-alphabet inputs has drawn increasing research interest in recent years [17]–[28].

As shown in Fig. 1, we consider the underlay cognitive multiple-access wiretap channel (CMAC-WT), where two secondary-user transmitters (STs) communicate with one secondary-user receiver (SR) in the presence of an ED and subject to interference threshold constraints at PRs. Each node in the system is equipped with multiple antennas. To the best of our knowledge, this is a general model that has not been addressed yet. We design linear precoding matrices to achieve the maximum average secrecy sum rate under finite-alphabet inputs and statistical CSI at STs. The problem setting is much closer to practical systems because it targets finite-alphabet inputs directly and exploits statistical CSI of fading channels. However, this problem is extremely difficult to solve due to two reasons: First, the computational complexity for evaluating the average secrecy sum rate is prohibitively high. Second, and more importantly, the optimization problem itself is a nonconvex and nondeterministic polynomial-time NP-hard problem.

A subset of nonconvex optimization, which is called the difference of convex functions (DC) optimization, has been extensively studied by exploiting its underlying structure [29]–[31]. DC optimization aims to maximize a DC function under some DC constraints. In [29], a basic outer-approximation framework was proposed for solving DC problems. In [30], a new DC algorithm was introduced by exploiting the duality theory of DC optimization. In [31], Yuille and Rangarajan presented the convex–concave procedure, which can be regarded as a special case of the algorithm in [30]. Since any twice continuously differentiable function is a DC function [29], our linear precoding problem is a DC optimization problem. However, no practical algorithm is known to construct DC decomposition for an arbitrary twice continuously differentiable function. Moreover, if we do not carefully design the DC representation of the average secrecy sum rate, the algorithms in [29]–[31] will suffer from very slow convergence [32]. Therefore, DC representation is a main factor that affects the performance of DC algorithms.

We solve our problem efficiently by combining the convex–concave procedure with an outer-approximation framework. We first exploit an accurate approximation of the average secrecy sum rate to reduce the complexity and then reformulate the approximated average secrecy sum rate as a DC function. Subsequently, we generate a sequence of relaxed sets, which can be explicitly expressed as the union of convex sets, to approach the nonconvex feasible set. This way, near-optimal precoders are obtained by maximizing the approximated average secrecy sum rate over these convex sets. Numerical results show that when considering finite-alphabet inputs, our proposed algorithm significantly outperforms the conventional Gaussian precoding method, which designs precoding matrices to maximize the average secrecy sum rate under Gaussian inputs, in the medium and high signal-to-noise ratio (SNR) regimes.

The rest of this paper is organized as follows. Section II introduces the system model and formulates the linear precoding problem, Section III develops a numerical algorithm to maximize the average secrecy sum rate under finite-alphabet inputs and statistical CSI, Section IV presents several numerical results, and Section V draws the conclusion.

Notations: Boldface lowercase letters, boldface uppercase letters, and calligraphic letters are used to denote vectors, matrices, and sets, respectively. The superscripts $(\cdot)^T$ and $(\cdot)^H$ represent transpose and Hermitian operations, respectively. $[\cdot]^+$ denotes $\max(\cdot, 0)$; $\mathrm{diag}(\cdot)$ represents a block diagonal matrix whose diagonal elements are matrices. $\mathrm{tr}(\cdot)$ is the trace of a matrix; $\mathrm{vec}(\cdot)$ is a column vector formed by stacking the columns of a matrix; $\|\cdot\|$ denotes the Euclidean norm of a vector; $\mathbf{A} \otimes \mathbf{B}$ is the Kronecker product of two matrices $\mathbf{A}$ and $\mathbf{B}$; $E(\cdot)$ represents the statistical expectation; $\Re(\cdot)$ and $\Im(\cdot)$ denote the real and imaginary parts of a complex vector or matrix; $\geq$ and $\leq$ are defined component-wise. $\mathbf{I}$ and $\mathbf{0}$ denote an identity matrix and a zero matrix, respectively, with appropriate dimensions; $\mathbf{A} \succeq \mathbf{0}$ denotes the positive semidefiniteness of $\mathbf{A}$. The symbol $\mathcal{I}(\cdot)$ represents the mutual information; $\log(\cdot)$ and $\ln(\cdot)$ are used for the base-2 logarithm and natural logarithm, respectively.

II. System Model and Problem Formulation

We consider the fading CMAC-WT shown in Fig. 1. The ith ST has N_{T_i} antennas, $i = 1, 2$; the SR has N_R antennas;

the ED has N_E antennas; and the jth PR has N_j antennas $j = 1, 2, \ldots, J$. The channel output at the SR, the ED, and the jth PR are, respectively, given by

$$\begin{aligned}\mathbf{y}_R &= \mathbf{H}_1\mathbf{P}_1\mathbf{s}_1 + \mathbf{H}_2\mathbf{P}_2\mathbf{s}_2 + \mathbf{n}_R\\ \mathbf{z}_E &= \mathbf{G}_1\mathbf{P}_1\mathbf{s}_1 + \mathbf{G}_2\mathbf{P}_2\mathbf{s}_2 + \mathbf{n}_E\\ \mathbf{w}_j &= \mathbf{F}_{1,j}\mathbf{P}_1\mathbf{s}_1 + \mathbf{F}_{2,j}\mathbf{P}_2\mathbf{s}_2 + \mathbf{n}_j,\ j = 1, 2, \ldots, J \end{aligned}\tag{1}$$

where $\mathbf{H}_i$, $\mathbf{G}_i$, and $\mathbf{F}_{i,j}$ are complex channel matrices from the ith ST to the SR, the ED, and the jth PR, respectively; $\mathbf{P}_i$ is the linear precoding matrix at the ith ST, $i = 1, 2$; $\mathbf{s}_i$ is the input data vector at the ith ST with zero mean and covariance $E_{\mathbf{s}_i}[\mathbf{s}_i\mathbf{s}_i^H] = \mathbf{I}$, $i = 1, 2$; and $\mathbf{n}_R$, $\mathbf{n}_E$, and $\mathbf{n}_j$ are independent and identically distributed (i.i.d.) zero-mean circularly symmetric complex Gaussian noise with covariance matrices $\sigma_R^2\mathbf{I}$, $\sigma_E^2\mathbf{I}$, and $\sigma_j^2\mathbf{I}$, respectively.

The channel matrices considered in this paper are modeled as [33]

$$\begin{aligned}\mathbf{H}_i &= \mathbf{\Phi}_h^{\frac{1}{2}}\tilde{\mathbf{H}}_i\mathbf{\Psi}_{h_i}^{\frac{1}{2}}, \quad i = 1, 2\\ \mathbf{G}_i &= \mathbf{\Phi}_g^{\frac{1}{2}}\tilde{\mathbf{G}}_i\mathbf{\Psi}_{g_i}^{\frac{1}{2}}, \quad i = 1, 2\\ \mathbf{F}_{i,j} &= \mathbf{\Phi}_{f_j}^{\frac{1}{2}}\tilde{\mathbf{F}}_{i,j}\mathbf{\Psi}_{f_{i,j}}^{\frac{1}{2}} \quad \forall(i,j)\end{aligned}\tag{2}$$

where $\tilde{\mathbf{H}}_i$, $\tilde{\mathbf{G}}_i$, and $\tilde{\mathbf{F}}_{i,j}$ are random matrices with i.i.d. zero-mean unit variance complex Gaussian entries; $\mathbf{\Phi}_h$, $\mathbf{\Phi}_g$, and $\mathbf{\Phi}_{f_j}$ are positive semidefinite receive correlation matrices of $\mathbf{H}_i$, $\mathbf{G}_i$, and $\mathbf{F}_{i,j}$, respectively; and $\mathbf{\Psi}_{h_i}$, $\mathbf{\Psi}_{g_i}$, and $\mathbf{\Psi}_{f_{i,j}}$ are positive semidefinite transmit correlation matrices of $\mathbf{H}_i$, $\mathbf{G}_i$, and $\mathbf{F}_{i,j}$, respectively.

We assume that the SR has instantaneous channel realizations of $\{\mathbf{H}_1, \mathbf{H}_2\}$, the ED has instantaneous channel realizations of $\{\mathbf{G}_1, \mathbf{G}_2\}$, and STs only know the transmit and receive correlation matrices of $\{\mathbf{H}_1, \mathbf{H}_2, \mathbf{G}_1, \mathbf{G}_2, \mathbf{F}_{i,j},\ \forall(i,j)\}$ as well as the distributions of $\{\tilde{\mathbf{H}}_1, \tilde{\mathbf{H}}_2, \tilde{\mathbf{G}}_1, \tilde{\mathbf{G}}_2, \tilde{\mathbf{F}}_{i,j},\ \forall(i,j)\}$. Under these assumptions, the following secrecy sum rate is achievable [9]:

$$\begin{aligned}&[\mathcal{I}(\mathbf{s}_1, \mathbf{s}_2; \mathbf{y}_R|\mathbf{H}) - \mathcal{I}(\mathbf{s}_1, \mathbf{s}_2; \mathbf{z}_E|\mathbf{G})]^+\\ &= \left[E_{\mathbf{H}}\mathcal{I}(\mathbf{s}_1, \mathbf{s}_2; \mathbf{y}_R|\mathbf{H} = \bar{\mathbf{H}}) - E_{\mathbf{G}}\mathcal{I}(\mathbf{s}_1, \mathbf{s}_2; \mathbf{z}_E|\mathbf{G} = \bar{\mathbf{G}})\right]^+\end{aligned}$$

where $\mathbf{H} = [\mathbf{H}_1, \mathbf{H}_2]$, and $\mathbf{G} = [\mathbf{G}_1, \mathbf{G}_2]$; $\bar{\mathbf{H}}$ and $\bar{\mathbf{G}}$ represent the instantaneous channel realizations of $\mathbf{H}$ and $\mathbf{G}$, respectively. For notational simplicity, we omit the given channel realization condition in mutual information expressions, and then, the average secrecy sum rate can be expressed as

$$R_{\text{avg}}(\mathbf{P}_1, \mathbf{P}_2) = [E_{\mathbf{H}}\mathcal{I}(\mathbf{s}_1, \mathbf{s}_2; \mathbf{y}_R) - E_{\mathbf{G}}\mathcal{I}(\mathbf{s}_1, \mathbf{s}_2; \mathbf{z}_E)]^+. \tag{3}$$

We maximize $R_{\text{avg}}(\mathbf{P}_1, \mathbf{P}_2)$ subject to power constraints at STs and interference threshold constraints at PRs. The average transmit power conforms to the power constraint β_i, i.e.,

$$E_{\mathbf{s}_i}\,\mathrm{tr}\left(\mathbf{P}_i\mathbf{s}_i\mathbf{s}_i^H\mathbf{P}_i^H\right) = \mathrm{tr}\left(\mathbf{P}_i^H\mathbf{P}_i\right) \le \beta_i, \quad i = 1, 2 \tag{4}$$

and the average interference power at the jth PR is limited by γ_j, i.e.,

$$\begin{aligned}&\sum_{i=1}^{2} E_{\mathbf{s}_i,\mathbf{F}_{i,j}}\left[\mathrm{tr}\left(\mathbf{F}_{i,j}\mathbf{P}_i\mathbf{s}_i\mathbf{s}_i^H\mathbf{P}_i^H\mathbf{F}_{i,j}^H\right)\right]\\ &\quad= \sum_{i=1}^{2} E_{\tilde{\mathbf{F}}_{i,j}}\left[\mathrm{tr}\left(\mathbf{P}_i^H\left(\mathbf{\Psi}_{f_{i,j}}^{\frac{1}{2}}\right)^H\tilde{\mathbf{F}}_{i,j}^H\mathbf{\Phi}_{f_j}\tilde{\mathbf{F}}_{i,j}\mathbf{\Psi}_{f_{i,j}}^{\frac{1}{2}}\mathbf{P}_i\right)\right]\\ &\quad= \mathrm{tr}\left(\mathbf{\Phi}_{f_j}\right)\cdot\sum_{i=1}^{2}\mathrm{tr}\left(\mathbf{P}_i^H\mathbf{\Psi}_{f_{i,j}}\mathbf{P}_i\right) \le \gamma_j \quad \forall j.\end{aligned}\tag{5}$$

The second equality in (5) holds because each element of $\tilde{\mathbf{F}}_{i,j}$ is an i.i.d. complex Gaussian variable with zero mean and unit variance, and $\tilde{\mathbf{F}}_{i,j}$ is independent of $\mathbf{s}_i$. Then, the average secrecy sum-rate maximization problem is formulated as

$$\begin{aligned}&\underset{\mathbf{P}_1,\mathbf{P}_2}{\text{maximize}} \quad R_{\text{avg}}(\mathbf{P}_1, \mathbf{P}_2)\\ &\text{subject to} \quad (4) \text{ and } (5).\end{aligned}\tag{6}$$

III. Linear Precoding Under Finite-Alphabet Inputs

Here, we solve problem (6) under finite-alphabet inputs. We assume that each symbol of the input data vector $\mathbf{s}_i$ is taken independently from an equiprobable discrete constellation with cardinality M_i, $i = 1, 2$. The average constellation-constrained mutual information $E_{\mathbf{H}}\mathcal{I}(\mathbf{s}_1, \mathbf{s}_2; \mathbf{y}_R)$ and $E_{\mathbf{G}}\mathcal{I}(\mathbf{s}_1, \mathbf{s}_2; \mathbf{z}_E)$ can then be expressed, respectively, as [19]

$$\begin{aligned}E_{\mathbf{H}}\mathcal{I}(\mathbf{s}; \mathbf{y}_R) = \log N - \frac{1}{N}\sum_{m=1}^{N} E_{\mathbf{H},\mathbf{n}_R}\\ \left\{\log\sum_{k=1}^{N}\exp\left(\frac{-\|\mathbf{H}\mathbf{P}\mathbf{e}_{mk} + \mathbf{n}_R\|^2 + \|\mathbf{n}_R\|^2}{\sigma_R^2}\right)\right\}\end{aligned}\tag{7}$$

$$\begin{aligned}E_{\mathbf{G}}\mathcal{I}(\mathbf{s}; \mathbf{z}_E) = \log N - \frac{1}{N}\sum_{m=1}^{N} E_{\mathbf{G},\mathbf{n}_E}\\ \left\{\log\sum_{k=1}^{N}\exp\left(\frac{-\|\mathbf{G}\mathbf{P}\mathbf{e}_{mk} + \mathbf{n}_E\|^2 + \|\mathbf{n}_E\|^2}{\sigma_E^2}\right)\right\}\end{aligned}\tag{8}$$

where $\mathbf{s} = [\mathbf{s}_1^T, \mathbf{s}_2^T]^T$, N is a constant and equal to $M_1^{N_{T_1}} M_2^{N_{T_2}}$, $\mathbf{P} = \mathrm{diag}(\mathbf{P}_1, \mathbf{P}_2)$, and $\mathbf{e}_{mk}$ is the difference between $\mathbf{d}_m$ and $\mathbf{d}_k$, with $\mathbf{d}_m$ and $\mathbf{d}_k$ representing two possible distinct signal vectors from $\mathbf{s}$.

Clearly, the evaluation and optimization of the given average mutual information is a difficult task. To obtain $E_{\mathbf{H}}\mathcal{I}(\mathbf{s}; \mathbf{y}_R)$ and $E_{\mathbf{G}}\mathcal{I}(\mathbf{s}; \mathbf{z}_E)$, we need to calculate expectations over $\mathbf{H}$ and $\mathbf{G}$ as well as $\mathbf{n}_R$ and $\mathbf{n}_E$. Unfortunately, these expectations have no closed-form expressions. Although we can use the Monte Carlo method to estimate these expectations, the computational complexity is prohibitively high particularly when the dimensions of $\mathbf{H}$ and $\mathbf{G}$ are large.

This difficulty can be mitigated by employing accurate approximations of (7) and (8). Based on [21], $E_{\mathbf{H}}\mathcal{I}(\mathbf{s};\mathbf{y}_R)$ and $E_{\mathbf{G}}\mathcal{I}(\mathbf{s};\mathbf{z}_E)$ can be approximated, respectively, as

$$\mathcal{I}_A(\mathbf{s};\mathbf{y}_R) = \log N - \frac{1}{N}\sum_{m=1}^{N}\log\sum_{k=1}^{N} \prod_q \left(1+\frac{h_q}{2\sigma_R^2}\mathbf{e}_{mk}^H\mathbf{P}^H\boldsymbol{\Psi}_h\mathbf{P}\mathbf{e}_{mk}\right)^{-1} \tag{9}$$

$$\mathcal{I}_A(\mathbf{s};\mathbf{z}_E) = \log N - \frac{1}{N}\sum_{m=1}^{N}\log\sum_{k=1}^{N} \prod_q \left(1+\frac{g_q}{2\sigma_E^2}\mathbf{e}_{mk}^H\mathbf{P}^H\boldsymbol{\Psi}_g\mathbf{P}\mathbf{e}_{mk}\right)^{-1} \tag{10}$$

where $\boldsymbol{\Psi}_h = \mathrm{diag}(\boldsymbol{\Psi}_{h_1},\boldsymbol{\Psi}_{h_2})$, and $\boldsymbol{\Psi}_g = \mathrm{diag}(\boldsymbol{\Psi}_{g_1},\boldsymbol{\Psi}_{g_2})$; h_q and g_q represent the qth eigenvalue of $\boldsymbol{\Phi}_h$ and $\boldsymbol{\Phi}_g$, respectively. Approximations (9) and (10) are very accurate for arbitrary correlation matrices and precoders, and the computational complexity of (9) and (10) is several orders of magnitude lower than that of the original average mutual information [21].

By replacing $R_{\mathrm{avg}}(\mathbf{P}_1,\mathbf{P}_2)$ with $[\mathcal{I}_A(\mathbf{s};\mathbf{y}_R)-\mathcal{I}_A(\mathbf{s};\mathbf{z}_E)]^+$, problem (6) can be approximated as

$$\begin{aligned}&\underset{\mathbf{P}_1,\mathbf{P}_2}{\text{maximize}} \quad [\mathcal{I}_A(\mathbf{s};\mathbf{y}_R)-\mathcal{I}_A(\mathbf{s};\mathbf{z}_E)]^+\\ &\text{subject to} \quad (4) \text{ and } (5).\end{aligned} \tag{11}$$

A. Precoder Vectorization

We reformulate (11) into a vectorized form by employing the precoder vectorization technique [28], [34]. This reformulation can better exploit the inherent structure of (11). For convenience, we first reformulate $\mathcal{I}_A(\mathbf{s};\mathbf{y}_R)$ by precoder vectorization, and then, the same procedure can be applied for $\mathcal{I}_A(\mathbf{s};\mathbf{z}_E)$ and the constraints of problem (11).

We start by rewriting $\mathbf{e}_{mk}^H\mathbf{P}^H\boldsymbol{\Psi}_h\mathbf{P}\mathbf{e}_{mk}$ as

$$\mathbf{e}_{mk}^H\mathbf{P}^H\boldsymbol{\Psi}_h\mathbf{P}\mathbf{e}_{mk} = \sum_{i=1}^{2}\mathbf{e}_{mk,i}^H\mathbf{P}_i^H\boldsymbol{\Psi}_{h_i}\mathbf{P}_i\mathbf{e}_{mk,i} \tag{12}$$

where $\mathbf{e}_{mk} = [\mathbf{e}_{mk,1}^T,\mathbf{e}_{mk,2}^T]^T$. Using the following matrix equation [35]:

$$\mathrm{tr}(\mathbf{A}^T\mathbf{B}\mathbf{A}\mathbf{C}) = \mathrm{vec}(\mathbf{A})^T\cdot(\mathbf{C}^T\otimes\mathbf{B})\cdot\mathrm{vec}(\mathbf{A}) \tag{13}$$

$\mathbf{e}_{mk,i}^H\mathbf{P}_i^H\boldsymbol{\Psi}_{h_i}\mathbf{P}_i\mathbf{e}_{mk,i}$ can be rewritten as

$$\begin{aligned}\mathbf{e}_{mk,i}^H\mathbf{P}_i^H\boldsymbol{\Psi}_{h_i}\mathbf{P}_i\mathbf{e}_{mk,i} &= \mathrm{tr}\left(\mathbf{P}_i^H\boldsymbol{\Psi}_{h_i}\mathbf{P}_i\mathbf{E}_{mk,i}^T\right)\\ &= \mathrm{vec}(\mathbf{P}_i)^H\cdot(\mathbf{E}_{mk,i}\otimes\boldsymbol{\Psi}_{h_i})\cdot\mathrm{vec}(\mathbf{P}_i)\end{aligned} \tag{14}$$

where $\mathbf{E}_{mk,i} = (\mathbf{e}_{mk,i}\mathbf{e}_{mk,i}^H)^T$. By letting

$$\hat{\mathbf{p}} = \begin{bmatrix}\mathrm{vec}(\mathbf{P}_1)\\ \mathrm{vec}(\mathbf{P}_2)\end{bmatrix}, \quad \mathbf{p} = \begin{bmatrix}\Re\{\hat{\mathbf{p}}\}\\ \Im\{\hat{\mathbf{p}}\}\end{bmatrix} \tag{15}$$

$$\hat{\mathbf{A}}_{mk} = \frac{1}{2}\cdot\mathrm{diag}\left(\mathbf{E}_{mk,1}\otimes\boldsymbol{\Psi}_{h_1},\mathbf{E}_{mk,2}\otimes\boldsymbol{\Psi}_{h_2}\right) \tag{16}$$

$$\mathbf{A}_{mk} = \begin{bmatrix}\Re\{\hat{\mathbf{A}}_{mk}\} & -\Im\{\hat{\mathbf{A}}_{mk}\}\\ \Im\{\hat{\mathbf{A}}_{mk}\} & \Re\{\hat{\mathbf{A}}_{mk}\}\end{bmatrix} \tag{17}$$

$\mathcal{I}_A(\mathbf{s};\mathbf{y}_R)$ can be expressed alternatively as

$$\mathcal{I}_A(\mathbf{s};\mathbf{y}_R) = \log N - \frac{1}{N}\sum_{m=1}^{N}\log\sum_{k=1}^{N} \prod_q \left(1+\frac{h_q}{\sigma_R^2}\cdot\mathbf{p}^T\mathbf{A}_{mk}\mathbf{p}\right)^{-1}. \tag{18}$$

Here, $\mathbf{A}_{mk}\succeq\mathbf{0}$ because $\mathbf{p}^T\mathbf{A}_{mk}\mathbf{p}$ is equal to $\|\boldsymbol{\Psi}_h^{1/2}\mathbf{P}\mathbf{e}_{mk}\|^2$, which is nonnegative.

Similarly, we define $\hat{\mathbf{B}}_{mk}$ and $\mathbf{B}_{mk}$ as

$$\hat{\mathbf{B}}_{mk} = \frac{1}{2}\cdot\mathrm{diag}\left(\mathbf{E}_{mk,1}\otimes\boldsymbol{\Psi}_{g_1},\mathbf{E}_{mk,2}\otimes\boldsymbol{\Psi}_{g_2}\right) \tag{19}$$

$$\mathbf{B}_{mk} = \begin{bmatrix}\Re\{\hat{\mathbf{B}}_{mk}\} & -\Im\{\hat{\mathbf{B}}_{mk}\}\\ \Im\{\hat{\mathbf{B}}_{mk}\} & \Re\{\hat{\mathbf{B}}_{mk}\}\end{bmatrix}\succeq\mathbf{0} \tag{20}$$

$\hat{\mathbf{C}}_i$ and $\mathbf{C}_i$ as

$$\hat{\mathbf{C}}_i = \mathrm{diag}\left(\mathbf{I}\otimes(2-i)\cdot\mathbf{I},\mathbf{I}\otimes(i-1)\cdot\mathbf{I}\right) \tag{21}$$

$$\mathbf{C}_i = \begin{bmatrix}\Re\{\hat{\mathbf{C}}_i\} & -\Im\{\hat{\mathbf{C}}_i\}\\ \Im\{\hat{\mathbf{C}}_i\} & \Re\{\hat{\mathbf{C}}_i\}\end{bmatrix}\succeq\mathbf{0} \tag{22}$$

and $\hat{\mathbf{D}}_j$ and $\mathbf{D}_j$ as

$$\hat{\mathbf{D}}_j = \mathrm{tr}\left(\boldsymbol{\Phi}_{f_j}\right)\cdot\mathrm{diag}\left(\mathbf{I}\otimes\boldsymbol{\Psi}_{f_{1,j}},\mathbf{I}\otimes\boldsymbol{\Psi}_{f_{2,j}}\right) \tag{23}$$

$$\mathbf{D}_j = \begin{bmatrix}\Re\{\hat{\mathbf{D}}_j\} & -\Im\{\hat{\mathbf{D}}_j\}\\ \Im\{\hat{\mathbf{D}}_j\} & \Re\{\hat{\mathbf{D}}_j\}\end{bmatrix}\succeq\mathbf{0}. \tag{24}$$

Then, (11) is converted into a vectorized form, i.e.,

$$\underset{\mathbf{p}\in\mathcal{P}}{\text{maximize}} \quad [f(\mathbf{p})-g(\mathbf{p})]^+ \tag{25}$$

where $f(\mathbf{p})$ and $g(\mathbf{p})$ are given as

$$f(\mathbf{p}) = \frac{1}{N}\sum_{m=1}^{N}\log\sum_{k=1}^{N}\prod_q\left(1+\frac{g_q}{\sigma_E^2}\cdot\mathbf{p}^T\mathbf{B}_{mk}\mathbf{p}\right)^{-1} \tag{26}$$

$$g(\mathbf{p}) = \frac{1}{N}\sum_{m=1}^{N}\log\sum_{k=1}^{N}\prod_q\left(1+\frac{h_q}{\sigma_R^2}\cdot\mathbf{p}^T\mathbf{A}_{mk}\mathbf{p}\right)^{-1} \tag{27}$$

and $\mathcal{P}$ is the feasible set, i.e.,

$$\mathcal{P} = \{\mathbf{p}|\mathbf{p}^T\mathbf{C}_i\mathbf{p}\le\beta_i, i=1,2,\mathbf{p}^T\mathbf{D}_j\mathbf{p}\le\gamma_j\ \forall j\}. \tag{28}$$

The feasible set $\mathcal{P}$ is convex and compact because it can be geometrically interpreted as the intersection of multiple ellipsoids. The objective function $[f(\mathbf{p})-g(\mathbf{p})]^+$ is continuous over $\mathcal{P}$ because both $f(\mathbf{p})$ and $g(\mathbf{p})$ are continuous functions. Therefore, the existence of a globally optimal solution is guaranteed by the Weierstrass extreme value theorem [36]. In addition, the operator $[\cdot]^+$ has no effect on the optimal value of problem (25) and, thus, can be removed from the objective function because $\mathbf{p}=\mathbf{0}$ always belongs to $\mathcal{P}$. However, it is extremely difficult to solve problem (25) due to the following reasons: First, both $f(\mathbf{p})$ and $g(\mathbf{p})$ are neither convex nor concave; thus, (25) is a purely nonconvex optimization problem. Second, problem (25) is an NP-hard problem because a specialized problem with particular parameters $\mathbf{A}_{mk}$ and $\mathbf{B}_{mk}$ is NP-hard [37].

Although $f(\mathbf{p}) - g(\mathbf{p})$ is nonconcave, it can be expressed as a DC function by adding a convex term, i.e.,

$$\sigma(\mathbf{p}) = k \cdot \mathbf{p}^T \mathbf{p}, \ k > 0. \tag{29}$$

We can prove that both $f(\mathbf{p}) + \sigma(\mathbf{p})$ and $g(\mathbf{p}) + \sigma(\mathbf{p})$ are convex functions if

$$k \geq \alpha \cdot \max\left(\mathrm{tr}(\mathbf{\Phi}_h) \cdot \lambda_{\max}(\mathbf{\Psi}_h), \mathrm{tr}(\mathbf{\Phi}_g) \cdot \lambda_{\max}(\mathbf{\Psi}_g)\right) \tag{30}$$

where $\alpha = \sum_{m,k} \|\mathbf{e}_{mk}\|^2$, and $\lambda_{\max}(\cdot)$ represents the maximum eigenvalue of a matrix. Then, $[f(\mathbf{p}) + \sigma(\mathbf{p})] - [g(\mathbf{p}) + \sigma(\mathbf{p})]$ is an explicit DC function, and problem (25) can be solved by DC algorithms. However, this DC representation is not efficient because k is too large [32]. Through extensive simulations, we observe that even when each node in the system is only equipped with two antennas, the DC algorithm with this representation cannot converge within hundreds of thousands of iterations. Therefore, a computationally efficient DC representation of the approximated average secrecy sum rate is crucial for designing our algorithm.

B. Outer Approximation of the Feasible Set

We first rewrite (25) with an additional hyperrectangle $\mathcal{B}_{\text{init}}$, i.e.,

$$\underset{\mathbf{p} \in \mathcal{P} \cap \mathcal{B}_{\text{init}}}{\text{maximize}} \quad f(\mathbf{p}) - g(\mathbf{p}) \tag{31}$$

in which the hyperrectangle $\mathcal{B}_{\text{init}}$ is given by

$$\mathcal{B}_{\text{init}} = \left\{ \mathbf{p} | \mathbf{l}(\mathcal{B}_{\text{init}}) \leq \mathbf{p} \leq \mathbf{u}(\mathcal{B}_{\text{init}}) \right\}. \tag{32}$$

To ensure that problems (31) and (25) are equivalent, the hyperrectangle $\mathcal{B}_{\text{init}}$ should contain the feasible set $\mathcal{P}$, i.e., $\mathcal{P} \subseteq \mathcal{B}_{\text{init}}$. Let u_i and l_i denote the ith component of $\mathbf{u}(\mathcal{B}_{\text{init}})$ and $\mathbf{l}(\mathcal{B}_{\text{init}})$, respectively. $\mathcal{B}_{\text{init}}$ can be obtained via solving the following concave maximization problem:

$$u_i = \underset{\mathbf{p} \in \mathcal{P}}{\text{maximize}} \quad p_i \tag{33}$$

where p_i is the ith component of $\mathbf{p}$. Due to the symmetry of problem (33), l_i can be set as $-u_i$.

By introducing a new variable $\mathbf{Q} = \mathbf{p}\mathbf{p}^T$, we define a set function $\varphi(\mathcal{F})$ as the optimal value of the following optimization problem:

$$\varphi(\mathcal{F}) \triangleq \underset{(\mathbf{Q},\mathbf{p}) \in \mathcal{F}}{\text{maximize}} \quad F(\mathbf{Q}) - G(\mathbf{Q}) \tag{34}$$

where $F(\mathbf{Q})$ and $G(\mathbf{Q})$ are given as

$$F(\mathbf{Q}) = \frac{1}{N} \sum_{m=1}^{N} \log \sum_{k=1}^{N} \prod_{q} \left(1 + \frac{g_q}{\sigma_E^2} \cdot \mathrm{tr}(\mathbf{B}_{mk}\mathbf{Q})\right)^{-1} \tag{35}$$

$$G(\mathbf{Q}) = \frac{1}{N} \sum_{m=1}^{N} \log \sum_{k=1}^{N} \prod_{q} \left(1 + \frac{h_q}{\sigma_R^2} \cdot \mathrm{tr}(\mathbf{A}_{mk}\mathbf{Q})\right)^{-1}. \tag{36}$$

Note that $F(\mathbf{Q})$ and $G(\mathbf{Q})$ are convex functions because 1) $\log \sum_k \prod_q f_{q,k}^{-1}$ can be written as $\log \sum_k \exp(-\sum_q \ln f_{q,k})$, and 2) $\log \sum_k \exp(g_k)$ is convex whenever the g_k values are convex [38]. Therefore, $F(\mathbf{Q}) - G(\mathbf{Q})$ is a DC function. Furthermore, when $\mathcal{F}_{\text{init}}$, which is given by

$$\mathcal{F}_{\text{init}} = \left\{ (\mathbf{Q}, \mathbf{p}) \middle| \begin{array}{l} \mathbf{Q} = \mathbf{p}\mathbf{p}^T, \mathrm{tr}(\mathbf{D}_j\mathbf{Q}) \leq \gamma_j \ \forall j \\ \mathbf{p} \in \mathcal{B}_{\text{init}}, \mathrm{tr}(\mathbf{C}_i\mathbf{Q}) \leq \beta_i, i = 1{,}2 \end{array} \right\} \tag{37}$$

is equivalent to the feasible set $\mathcal{P}$, $\varphi(\mathcal{F}_{\text{init}})$ serves as the optimal value of problem (31). However, it is very difficult to obtain $\varphi(\mathcal{F}_{\text{init}})$ directly because $\mathcal{F}_{\text{init}}$ is a nonconvex set. Although we can use semidefinite relaxation (SDR) to relax $\mathcal{F}_{\text{init}}$ into a convex set by relaxing the nonconvex part $\mathbf{Q} = \mathbf{p}\mathbf{p}^T$, the solution obtained by SDR is not optimal and cannot be iteratively improved. Hence, we need tighter relaxations to overcome the shortcomings of SDR.

The key idea of our proposed precoding algorithm is to generate a sequence of asymptotically tight sets $\{\mathcal{F}_k\}$ to approach $\mathcal{F}_{\text{init}}$, and then, $\varphi(\mathcal{F}_{\text{init}})$ can be iteratively approached from above by solving a sequence of optimization problems $\{\varphi(\mathcal{F}_k)\}$. The sequence $\{\mathcal{F}_k\}$ should satisfy the following three properties:

$$\begin{gathered} \mathcal{F}_1 \supseteq \mathcal{F}_2 \supseteq \cdots \supseteq \mathcal{F}_{\text{init}} \\ \lim_{k \to \infty} \varphi(\mathcal{F}_k) = \varphi(\mathcal{F}_{\text{init}}) \\ \mathcal{F}_k = \bigcup_{i=1}^{k} \mathcal{C}(\mathcal{B}_i) \quad \forall k \end{gathered} \tag{38}$$

where $\mathcal{C}(\mathcal{B}_i)$ is a convex set to be defined in (40). The first property implies that $\{\varphi(\mathcal{F}_k)\}$ is a monotonically decreasing sequence bounded below by $\varphi(\mathcal{F}_{\text{init}})$. The second property guarantees that $\varphi(\mathcal{F}_{\text{init}})$ can be readily obtained by the sequence $\{\varphi(\mathcal{F}_k)\}$. The last property provides a tractable way to compute $\{\varphi(\mathcal{F}_k)\}$, i.e.,

$$\varphi(\mathcal{F}_k) = \max_{1 \leq i \leq k} \varphi\left(\mathcal{C}(\mathcal{B}_i)\right). \tag{39}$$

Based on (38), achieving $\varphi(\mathcal{F}_{\text{init}})$ may need a sufficiently large number of iterations, which is not practical when the computational time is concerned. To address this issue, we also generate a lower bound of $\varphi(\mathcal{F}_{\text{init}})$ in each iteration. Denote the optimal solution for $\varphi(\mathcal{F}_k)$ at the kth iteration by $(\mathbf{Q}_k^{\text{opt}}, \mathbf{p}_k^{\text{opt}})$. We extract a feasible solution of problem (31) from $\mathbf{Q}_k^{\text{opt}}$, and the corresponding approximated average secrecy sum rate is denoted by $\varphi_L(\mathcal{F}_k)$, which serves as a lower bound of $\varphi(\mathcal{F}_{\text{init}})$.

In the remaining part of this section, we construct $\{\mathcal{F}_k\}$ explicitly as the union of convex sets $\{\mathcal{C}(\mathcal{B}_i)\}$. The approximated average secrecy sum-rate maximization problem over $\mathcal{C}(\mathcal{B}_i)$ and an efficient method to generate the lower bound $\varphi_L(\mathcal{F}_k)$ are investigated in the following section.

For ease of exposition, we first define a convex set $\mathcal{C}(\mathcal{B})$ as

$$\mathcal{C}(\mathcal{B}) \triangleq \left\{ (\mathbf{Q}, \mathbf{p}) \middle| \begin{array}{l} \mathbf{Q} \succeq \mathbf{p}\mathbf{p}^T, \mathrm{tr}(\mathbf{C}_i\mathbf{Q}) \leq \beta_i, i = 1, 2 \\ (\mathbf{Q}, \mathbf{p}) \in \mathcal{S}(\mathcal{B}), \mathrm{tr}(\mathbf{D}_j\mathbf{Q}) \leq \gamma_j \ \forall j \end{array} \right\} \tag{40}$$

where $\mathcal{S}(\mathcal{B})$ is another convex set given by

$$\mathcal{S}(\mathcal{B}) \triangleq \left\{ (\mathbf{Q}, \mathbf{p}) \middle| \begin{array}{l} \mathbf{Q} - \mathbf{L_p} - \mathbf{L_p}^T + \mathbf{l}(\mathcal{B}) \cdot \mathbf{l}(\mathcal{B})^T \geq \mathbf{0} \\ \mathbf{Q} - \mathbf{U_p} - \mathbf{U_p}^T + \mathbf{u}(\mathcal{B}) \cdot \mathbf{u}(\mathcal{B})^T \geq \mathbf{0} \\ \mathbf{Q} - \mathbf{L_p} - \mathbf{U_p}^T + \mathbf{l}(\mathcal{B}) \cdot \mathbf{u}(\mathcal{B})^T \geq \mathbf{0} \\ \mathbf{l}(\mathcal{B}) \leq \mathbf{p} \leq \mathbf{u}(\mathcal{B}) \end{array} \right\} \tag{41}$$

with $\mathbf{L_p} = \mathbf{l}(\mathcal{B}) \cdot \mathbf{p}^T$ and $\mathbf{U_p} = \mathbf{u}(\mathcal{B}) \cdot \mathbf{p}^T$. The following two propositions are the foundation for constructing $\{\mathcal{F}_k\}$.

Proposition 1: If we split the initial hyperrectangle $\mathcal{B}_{\text{init}}$ into K smaller hyperrectangles such that $\mathcal{B}_{\text{init}} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_K$, then $\mathcal{F}_{\text{init}} \subseteq \mathcal{C}(\mathcal{B}_1) \cup \cdots \cup \mathcal{C}(\mathcal{B}_K)$.

Proof: See Appendix A. ■

Proposition 2: If we split a hyperrectangle $\mathcal{B}$ into two smaller hyperrectangles $\mathcal{B}_1$ and $\mathcal{B}_2$ such that $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ and $\mathcal{B}_1 \cap \mathcal{B}_2 = \emptyset$, then $\mathcal{C}(\mathcal{B}_1) \cup \mathcal{C}(\mathcal{B}_2) \subseteq \mathcal{C}(\mathcal{B})$.

Proof: See Appendix A. ■

With the help of Proposition 1, the first relaxed set $\mathcal{F}_1$ is obtained as

$$\mathcal{F}_1 = \mathcal{C}(\mathcal{B}_{\text{init}}). \tag{42}$$

Similarly, in the second iteration, we generate $\mathcal{F}_2$ by partitioning the initial hyperrectangle $\mathcal{B}_{\text{init}}$ into two nonintersection hyperrectangles $\mathcal{B}_1$ and $\mathcal{B}_2$, i.e.,

$$\mathcal{F}_2 = \mathcal{C}(\mathcal{B}_1) \cup \mathcal{C}(\mathcal{B}_2) \subseteq \mathcal{F}_1. \tag{43}$$

We continue this process to generate a sequence of relaxed sets $\{\mathcal{F}_k\}$ satisfying (38). At the kth iteration, $\mathcal{B}_{\text{init}}$ is split into k nonintersection hyperrectangles $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_k$ such that

$$\mathcal{F}_k = \mathcal{C}(\mathcal{B}_1) \cup \cdots \cup \mathcal{C}(\mathcal{B}_k). \tag{44}$$

The outer-approximation algorithm is summarized in Algorithm 1.

Algorithm 1 The outer-approximation algorithm

```
1) Initialization: Set the maximum number of iterations
   $K_{\max}$, $k = 1$, $\mathbb{B} = \{\mathcal{B}_{\text{init}}\}$, $\mathcal{F}_1 = \mathcal{C}(\mathcal{B}_{\text{init}})$, $U_1 = \varphi(\mathcal{F}_1)$,
   and $L_1 = \varphi_L(\mathcal{F}_1)$.
2) Stopping criterion: If $k \leq K_{\max}$, go to the next step;
   otherwise, STOP.
3) Partition criterion:
   a) select $\mathcal{B}_g = \arg\max_{\mathcal{B} \in \mathbb{B}} \{\varphi(\mathcal{C}(\mathcal{B}))\}$.
   b) split $\mathcal{B}_g$ along any of its longest edge into two small
      hyperrectangles, $\mathcal{B}_I$ and $\mathcal{B}_{II}$, with equal volume.
   c) remove $\mathcal{B}_g$ from $\mathbb{B}$, and add $\mathcal{B}_I$ and $\mathcal{B}_{II}$ into $\mathbb{B}$.
   d) compute the upper and lower bounds of $\varphi(\mathcal{F}_{\text{init}})$
         $\mathcal{F}_{k+1} = \bigcup_{\mathcal{B} \in \mathbb{B}} \mathcal{C}(\mathcal{B})$
         $U_{k+1} = \varphi(\mathcal{F}_{k+1}) = \max_{\mathcal{B} \in \mathbb{B}} \{\varphi(\mathcal{C}(\mathcal{B}))\}$
         $L_{k+1} = \varphi_L(\mathcal{F}_{k+1})$.
4) Set $k := k + 1$ and go to step 2).
```

The convergence of Algorithm 1 is presented by the following proposition.

Proposition 3: The sequence $\{\varphi(\mathcal{F}_k)\}$ converges to $\varphi(\mathcal{F}_{\text{init}})$, i.e., $\forall \varepsilon > 0$, $\exists K > 0$, such that $k > K$ implies $\varphi(\mathcal{F}_{\text{init}}) < \varphi(\mathcal{F}_k) < \varphi(\mathcal{F}_{\text{init}}) + \varepsilon$.

Proof: See Appendix B. ■

It is worth remarking that each relaxed set $\mathcal{F}_k$ is tighter than the set relaxed by SDR. We denote $\mathcal{F}_{\text{sdr}} = \{\mathbf{Q} | \mathbf{Q} \succeq \mathbf{0}, \text{tr}(\mathbf{C}_i\mathbf{Q}) \leq \beta_i, i = 1, 2, \text{tr}(\mathbf{D}_j\mathbf{Q}) \leq \gamma_j, \forall j\}$. Since $\mathbf{p}\mathbf{p}^T \succeq \mathbf{0}$, we have $\{\mathbf{Q} | (\mathbf{Q}, \mathbf{p}) \in \mathcal{F}_k\} \subseteq \mathcal{F}_{\text{sdr}}$ for any k. Thus, the solution obtained by Algorithm 1 is better than that obtained by the SDR method.

C. DC Optimization Over the Convex Set

Here, we maximize the approximated average sum rate over the convex set $\mathcal{C}(\mathcal{B})$ by employing the convex–concave procedure [31]. The convex–concave procedure is a general polynomial-time algorithm for solving DC problems, and it works quite well in practice [39]–[41]. We first rewrite the optimization problem as follows:

$$\varphi\left(\mathcal{C}(\mathcal{B})\right) = \underset{(\mathbf{Q}, \mathbf{p}) \in \mathcal{C}(\mathcal{B})}{\text{maximize}} \quad F(\mathbf{Q}) - G(\mathbf{Q}). \tag{45}$$

The objective function of problem (45) is a DC function, and the convex part $F(\mathbf{Q})$ can be lower bounded by its tangent at any point $\mathbf{Q}_c \succeq \mathbf{0}$, i.e.,

$$F(\mathbf{Q}) \geq F(\mathbf{Q}_c) + \text{tr}\left\{\nabla F(\mathbf{Q}_c)^T (\mathbf{Q} - \mathbf{Q}_c)\right\} \tag{46}$$

where $\nabla F(\mathbf{Q}_c)$ is the gradient of $F(\mathbf{Q})$ at $\mathbf{Q}_c$, i.e.,

$$\nabla F(\mathbf{Q}_c) = -\frac{1}{N} \sum_{m,k} w_{mk} \sum_{q} \frac{g_q \cdot \mathbf{B}_{mk}^T}{\sigma_E^2 + g_q \cdot \text{tr}(\mathbf{B}_{mk}\mathbf{Q}_c)} \tag{47}$$

with

$$w_{mk} = \frac{1}{\ln(2)} \cdot \frac{\exp\left\{\sigma_E^2 + g_q \cdot \text{tr}(\mathbf{B}_{mk}\mathbf{Q}_c)\right\}}{\sum_k \exp\left\{\sigma_E^2 + g_q \cdot \text{tr}(\mathbf{B}_{mk}\mathbf{Q}_c)\right\}}. \tag{48}$$

Therefore, by replacing $F(\mathbf{Q}) - G(\mathbf{Q})$ with a concave lower bound, i.e.,

$$\hat{F}(\mathbf{Q}; \mathbf{Q}_c) = F(\mathbf{Q}_c) + \text{tr}\left\{\nabla F(\mathbf{Q}_c)^T (\mathbf{Q} - \mathbf{Q}_c)\right\} - G(\mathbf{Q}) \tag{49}$$

we obtain the following concave maximization problem:

$$\underset{(\mathbf{Q}, \mathbf{p}) \in \mathcal{C}(\mathcal{B})}{\text{maximize}} \quad \hat{F}(\mathbf{Q}; \mathbf{Q}_c). \tag{50}$$

The convex–concave procedure obtains a locally optimal solution of problem (45) by solving a sequence of concave maximization problems (50) with different $\mathbf{Q}_c$ values. Once the optimal solution of (50) in the first iteration is found at initial $\mathbf{Q}_c$, which is denoted as $\mathbf{Q}_1^*$, the algorithm replaces $\mathbf{Q}_c$ with $\mathbf{Q}_1^*$ and then solve (50) again. At the nth iteration, the optimal solution of (50) is obtained by replacing $\mathbf{Q}_c$ with $\mathbf{Q}_{n-1}^*$, which is the optimal solution at the $(n-1)$th iteration. The convex–concave procedure for solving problem (45) is summarized in Algorithm 2.

Algorithm 2 The convex–concave procedure

```
1) Initialization: Given tolerance $\epsilon > 0$, choose a random
   initial point $\mathbf{Q}_0 \succeq \mathbf{0}$, set $N = 1$, $s_0 = F(\mathbf{Q}_0) - G(\mathbf{Q}_0)$,
   $s_1 = F(\mathbf{Q}_1^*) - G(\mathbf{Q}_1^*)$. Let $\mathbf{Q}_n^*$ represent the optimal so-
   lution of (50) at the $n$th iteration.
2) Stopping criterion: If $|s_n - s_{n-1}| > \epsilon$, go to the next step;
   otherwise, STOP.
3) Convex approximation:
   a) set $\mathbf{Q}_c = \mathbf{Q}_n^*$ and solve problem (50) to obtain $\mathbf{Q}_{n+1}^*$.
   b) set $s_{n+1} = F(\mathbf{Q}_{n+1}^*) - G(\mathbf{Q}_{n+1}^*)$ and $\mathbf{Q}_{\text{opt}} = \mathbf{Q}_{n+1}^*$.
4) Set $n := n + 1$ and go to step 2).
5) Output: $\mathbf{Q}_{\text{opt}}$ and $s_n$.
```

The stopping criterion in Algorithm 2 is guaranteed to be satisfied due to the following proposition.

Proposition 4: The sequence $\{s_n\}$ generated by Algorithm 2 is monotonically increasing, i.e., $s_{n+1} \geq s_n$.

Proof: Since the feasible set of (50) does not change in each iteration, the optimal solution in the nth iteration $\mathbf{Q}_n^*$ is a feasible point in the $(n+1)$th iteration. Thus, we have

$$\hat{F}\left(\mathbf{Q}_{n+1}^*; \mathbf{Q}_n^*\right) \geq \hat{F}\left(\mathbf{Q}_n^*; \mathbf{Q}_n^*\right) = s_n. \tag{51}$$

According to (46), it follows that

$$s_{n+1} = \hat{F}\left(\mathbf{Q}_{n+1}^*; \mathbf{Q}_{n+1}^*\right) \geq \hat{F}\left(\mathbf{Q}_{n+1}^*; \mathbf{Q}_n^*\right). \tag{52}$$

Therefore, $\{s_n\}$ is monotonically increasing. ■

Since problem (45) is nonconvex, Algorithm 2 is not guaranteed to converge to the globally optimal value $\varphi(\mathcal{C}(\mathcal{B}))$. Therefore, by embedding Algorithm 2 into Algorithm 1, we obtain a near-optimal solution $\mathbf{Q}_k^{\text{opt}}$ and the corresponding approximated upper bound of $\varphi(\mathcal{F}_{\text{init}})$ at the kth iteration of Algorithm 1. Simulation results show that the gap between the approximated upper bound and the actual upper bound is usually very small because Algorithm 2 is insensitive to the initial point $\mathbf{Q}_0$.

After obtaining $\mathbf{Q}_k^{\text{opt}}$ at the kth iteration, we need to get a feasible precoder pair $(\mathbf{P}_1, \mathbf{P}_2)$ and the corresponding lower bound $\varphi_L(\mathcal{F}_k)$. The feasible precoders can be obtained by extracting a feasible solution of (31) from $\mathbf{Q}_k^{\text{opt}}$. There are several rank-1 approximation methods to do this, and we adopt the Gaussian randomization procedure [42], which is summarized in Algorithm 3.

Algorithm 3 Gaussian randomization procedure

```
1) Given a number of randomizations $L$, and set $l = 1$.
2) If $l \leq L$, go to the next step; otherwise, STOP.
3) Generate $\boldsymbol{\xi}_l \sim N(\mathbf{0}, \mathbf{Q}_k^{\text{opt}})$, and construct a feasible point $\tilde{\mathbf{p}}_l$
   $$\tilde{\mathbf{p}}_l = \frac{\boldsymbol{\xi}_l}{\sqrt{\max\left\{\left\{\frac{\boldsymbol{\xi}_l^T\mathbf{C}_i\boldsymbol{\xi}_l}{\beta_i}\right\}_{i=1,2}, \left\{\frac{\boldsymbol{\xi}_l^T\mathbf{D}_j\boldsymbol{\xi}_l}{\gamma_j}\right\}_{\forall j}\right\}}}.$$
4) Set $l := l + 1$ and go to step 2).
5) Choose $\tilde{\mathbf{p}} = \arg\max_{1\leq l\leq L} \ f(\tilde{\mathbf{p}}_l) - g(\tilde{\mathbf{p}}_l)$.
6) Set $\varphi_L(\mathcal{F}_k) = f(\tilde{\mathbf{p}}) - g(\tilde{\mathbf{p}})$.
7) Recover $(\mathbf{P}_1, \mathbf{P}_2)$ from $\tilde{\mathbf{p}}$.
```

D. Complexity Analysis

The computational complexity of Algorithm 1 is analyzed as follows. In each iteration, Algorithm 1 invokes Algorithms 2 and 3 twice to calculate the approximated upper bound and the lower bound. Since the complexity of Algorithm 3 is negligible, the complexity order for Algorithm 1 is given by

$$2K_{\max} \cdot C \tag{53}$$

where $K_{\max}$ is the maximum number of iterations, and C is the complexity order for Algorithm 2. Algorithm 2 obtains the local maxima of problem (45) by solving a sequence of concave maximization problems (50). Each concave maximization problem (50) can be solved by the interior-point method, and the complexity order is about $\mathcal{O}(N^3)$ [38], where $N = 4(N_{T_1}^2 + N_{T_2}^2)^2 + 2(N_{T_1}^2 + N_{T_2}^2)$ is the total number of optimization variables in problem (50). Assuming that Algorithm 2 solves problems (50) by T times, the complexity order for Algorithm 2 is given by $\mathcal{O}(T \cdot N^3)$. Based on (53), the overall complexity order for Algorithm 1 is then $\mathcal{O}(2K_{\max}T \cdot N^3)$.

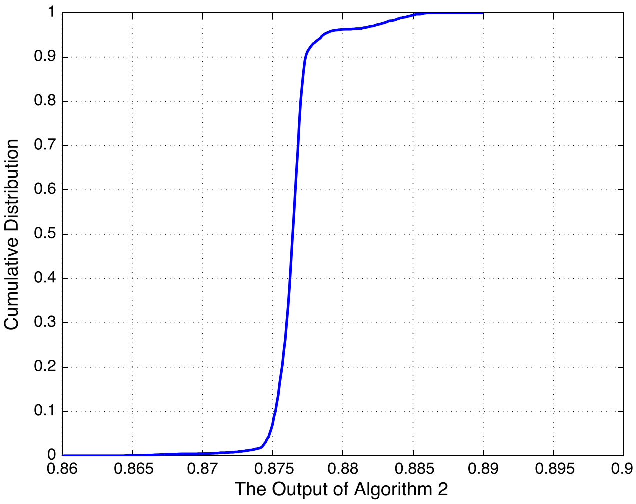


Fig. 2. Empirical cumulative distribution of the output s_n of Algorithm 2 from 3000 random initial points.

IV. Numerical Results

Here, we provide numerical results to demonstrate the efficacy of our proposed algorithm for the fading CMAC-WT under finite-alphabet inputs. For illustration purposes, we adopt the exponential correlation model, i.e.,

$$[\mathbf{C}(\rho)]_{i,j} = \rho^{|i-j|} \quad \forall(i,j) \tag{54}$$

where the scalar $\rho \in [0, 1)$ depicts the interference coupling between different antennas.

A. Convergence and Complexity Analysis

The convergence behavior of the proposed algorithm is demonstrated by considering a two-user fading CMAC-WT with two STs, one SR, one ED, and one PR. Each node in the system has two antennas. The correlation matrices are given by

$$\begin{aligned}\boldsymbol{\Phi}_h &= \mathbf{C}(0.3), \boldsymbol{\Psi}_{h_1} = \mathbf{C}(0.95), \boldsymbol{\Psi}_{h_2} = \mathbf{C}(0.85)\\ \boldsymbol{\Phi}_g &= \mathbf{C}(0.6), \boldsymbol{\Psi}_{g_1} = \mathbf{C}(0.4), \boldsymbol{\Psi}_{g_2} = \mathbf{C}(0.95)\\ \boldsymbol{\Phi}_f &= \mathbf{C}(0.5), \boldsymbol{\Psi}_{f_1} = \mathbf{C}(0.3), \boldsymbol{\Psi}_{f_2} = \mathbf{C}(0.5).\end{aligned} \tag{55}$$

The maximum transmit power is constrained by $\beta_1 = \beta_2 = 2$. The interference threshold is given as $\gamma = 0.2$. The input data vectors $\mathbf{s}_1$ and $\mathbf{s}_2$ are drawn independently from binary phase-shift keying (BPSK) constellation, and the noise power is set as $\sigma_R^2 = \sigma_E^2 = 0.1$.

The empirical cumulative distribution of the output s_n of Algorithm 2 from 3000 random initial points is shown in Fig. 2. The tolerance ε in Algorithm 2 is set as 0.002. $\mathbf{1}(\mathcal{B})$ and

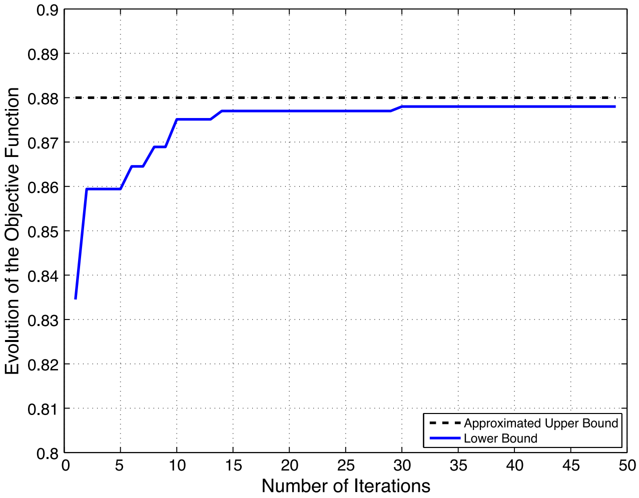


Fig. 3. Evolution of the objective function in (31) with BPSK inputs.

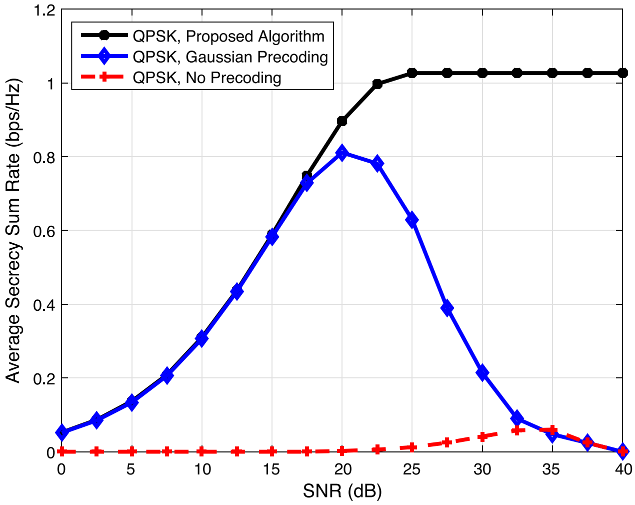


Fig. 4. Interference threshold at the PR is 10 dB less than the transmit power ($\gamma_1 = 0.2$).

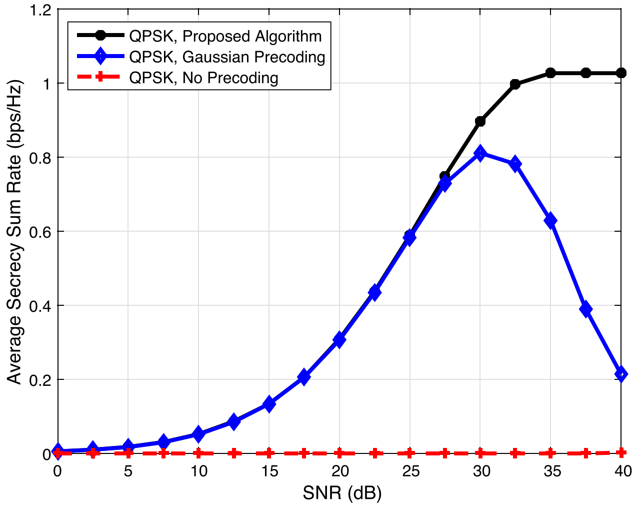


Fig. 5. Interference threshold at the PR is 20 dB less than the transmit power ($\gamma_1 = 0.02$).

$\mathbf{u}(\mathcal{B})$ are given as $\mathbf{l}(\mathcal{B}) = -\sqrt{2} \cdot \mathbf{1}$ and $\mathbf{u}(\mathcal{B}) = \sqrt{2} \cdot \mathbf{1}$. The empirical cumulative distribution illustrates that Algorithm 2 is insensitive to the initial point. Therefore, although problem (45) is nonconvex, the approximated upper bound obtained by Algorithm 2 is accurate.

Fig. 3 shows the evolution of the approximated upper bound and the lower bound of $\varphi(\mathcal{F}_{\text{init}})$. To guarantee that the approximated upper bound is accurate enough, the tolerance ε in Algorithm 2 is set as 0.001. In each iteration of Algorithm 1, we invoke Algorithm 2 to generate the approximated upper bound, which can be seen as the actual upper bound of $\varphi(\mathcal{F}_{\text{init}})$ according to the result in Fig. 2. We also invoke Algorithm 3 to generate feasible precoding matrices and the corresponding lower bound $\varphi_L(\mathcal{F}_{\text{init}})$. Note that when all hyperrectangles in $\mathbb{B}$ shrink down to a point, we can ensure that the approximated upper bound serves exactly as the actual upper bound. In the figure, we can see that after ten iterations, the gap between the approximated upper bound and the lower bound is less than 0.005. Moreover, near-optimal precoders within 0.002 tolerance are obtained through Algorithm 1 after 30 iterations.

B. Comparison With Other Possible Methods

Here, we consider a secure cognitive radio system that has two STs, one SR, one ED, and one PR. Each node in the system has two antennas. The correlation matrices are given by

$$\begin{aligned}
&\mathbf{\Phi}_h = \mathbf{C}(0.25), \mathbf{\Psi}_{h_1} = \mathbf{C}(0.95), \mathbf{\Psi}_{h_2} = \mathbf{C}(0.9)\\
&\mathbf{\Phi}_g = \mathbf{C}(0.75), \mathbf{\Psi}_{g_1} = \mathbf{C}(0.5), \mathbf{\Psi}_{g_2} = \mathbf{C}(0.3)\\
&\mathbf{\Phi}_f = \mathbf{C}(0.5), \mathbf{\Psi}_{f_1} = \mathbf{C}(0.8), \mathbf{\Psi}_{f_2} = \mathbf{C}(0.5).
\end{aligned} \tag{56}$$

The transmit power constraint is set as $\beta_1 = \beta_2 = \beta = 2$. The interference thresholds $\gamma_1 = 0.2$ and $\gamma_1 = 0.02$ are considered. The modulation is quaternary phase-shift keying (QPSK), and the noise variance $\sigma_R^2 = \sigma_E^2 = \sigma^2$. Then, the SNR can be defined as SNR $= \beta/\sigma^2$.

Figs. 4 and 5 show the comparison results between the Gaussian precoding method and the no-precoding case. The Gaussian precoding method is to design transmit covariance matrices that maximize the average secrecy sum rate under Gaussian signaling, i.e.,

$$\begin{aligned}
&\underset{\mathbf{Q}_1,\mathbf{Q}_2}{\text{maximize}} \quad E_{\mathbf{H}_1,\mathbf{H}_2}(R_1) - E_{\mathbf{G}_1,\mathbf{G}_2}(R_2)\\
&\text{subject to} \quad \text{tr}(\mathbf{Q}_i) \le \beta_i, \ i = 1, 2\\
&\qquad\qquad\quad \text{tr}\left(\mathbf{\Phi}_{f_j}\right) \cdot \text{tr}\left(\mathbf{Q}_1 \mathbf{\Psi}_{f_{1,j}} + \mathbf{Q}_2 \mathbf{\Psi}_{f_{2,j}}\right) \le \gamma_j \quad \forall j
\end{aligned} \tag{57}$$

where $\mathbf{Q}_i$ is the transmit covariance matrix of the ith ST, $i = 1, 2$; R_1 and R_2 are given by

$$R_1 = \log\det\left(\mathbf{I} + \frac{1}{\sigma_R^2}\mathbf{H}_1\mathbf{Q}_1\mathbf{H}_1^H + \frac{1}{\sigma_R^2}\mathbf{H}_2\mathbf{Q}_2\mathbf{H}_2^H\right) \tag{58}$$

$$R_2 = \log\det\left(\mathbf{I} + \frac{1}{\sigma_E^2}\mathbf{G}_1\mathbf{Q}_1\mathbf{G}_1^H + \frac{1}{\sigma_E^2}\mathbf{G}_2\mathbf{Q}_2\mathbf{G}_2^H\right). \tag{59}$$

Problem (57) is a DC optimization problem; thus, it can be efficiently solved by DC algorithms proposed in [41]. After obtaining the optimal transmit covariance matrices $(\bar{\mathbf{Q}}_1, \bar{\mathbf{Q}}_2)$, we can evaluate the finite-alphabet-based average secrecy sum rate under the corresponding optimal precoders $(\bar{\mathbf{Q}}_1^{1/2}, \bar{\mathbf{Q}}_2^{1/2})$. In the no-precoding case, we set precoding matrices as $\mathbf{P}_i = (\beta_i / N_{T_i})\mathbf{I}$, $i = 1, 2$, and then scale them down to meet interference threshold constraints, i.e.,

$$\bar{\mathbf{P}}_i = \left[\max_{1\le j\le J}\left\{\frac{\mathrm{tr}\left(\boldsymbol{\Phi}_{f_j}\right)}{\gamma_j}\cdot\sum_{i=1}^{2}\mathrm{tr}\left(\mathbf{P}_i^H\boldsymbol{\Psi}_{f_{i,j}}\mathbf{P}_i\right)\right\}\right]^{-\frac{1}{2}}\cdot\mathbf{P}_i. \quad (60)$$

Based on the results in Figs. 4 and 5, we have the following remarks.

1) In the low-SNR regime, our proposed precoding algorithm and the Gaussian precoding method have the same performance. According to [17], the low-SNR expansion of the mutual information is irrelevant to the input distribution; thus, the optimal precoders designed under Gaussian inputs are also optimal for the finite-alphabet input case.

2) In the medium- and high-SNR regimes, our proposed precoding algorithm offers a much higher average secrecy sum rate than the Gaussian precoding method. In Fig. 4, the normalized optimal precoders designed by our proposed precoding algorithm in the high-SNR regime are given by

$$\frac{1}{\sigma}\mathbf{P}_1^{\mathrm{opt}} = \begin{bmatrix} 0.663 + 0.008i & -1.188 + 0.277i \\ 0.663 + 0.008i & -1.188 + 0.277i \end{bmatrix} \quad (61)$$

$$\frac{1}{\sigma}\mathbf{P}_2^{\mathrm{opt}} = \begin{bmatrix} -0.578 + 0.399i & 1.209 - 0.459i \\ -0.578 + 0.399i & 1.209 - 0.459i \end{bmatrix}. \quad (62)$$

Equations (61) and (62) imply that when the noise power σ^2 is decreased, we should reduce the optimal transmit power $\mathrm{tr}((\mathbf{P}_1^{\mathrm{opt}})^H\mathbf{P}_1^{\mathrm{opt}})$ and $\mathrm{tr}((\mathbf{P}_2^{\mathrm{opt}})^H\mathbf{P}_2^{\mathrm{opt}})$ such that the average secrecy sum rate is kept at the maximum value of 1.0265 b/s/Hz in the high-SNR regime. Furthermore, the performance of the Gaussian precoding method severely degrades with the increasing SNR in the high-SNR regime because both $E_{\mathbf{H}}\mathcal{I}(\mathbf{s};\mathbf{y}_R)$ and $E_{\mathbf{G}}\mathcal{I}(\mathbf{s};\mathbf{z}_E)$ in (7) and (8) will saturate at $\log N$ in the high-SNR regime. Therefore, if we do not carefully control the transmit power, the average secrecy sum rate with finite-alphabet inputs $E_{\mathbf{H}}\mathcal{I}(\mathbf{s};\mathbf{y}_R) - E_{\mathbf{G}}\mathcal{I}(\mathbf{s};\mathbf{z}_E)$ will approach zero in the high-SNR regime. Since the Gaussian precoding method ignores the saturation property of finite-alphabet input systems, the corresponding average secrecy sum rate with finite-alphabet inputs severely degrades in the high-SNR regime.

3) Since the average secrecy sum rate for the Gaussian precoding method decreases with the increasing SNR in the high-SNR regime, we can use a portion of the available transmit power to make sure that the SNR is maintained at a certain level. The average secrecy sum rate is then kept at its maximum value. This simple power control method has been used in [11] and [12] to improve the secrecy sum-rate performance in the high-SNR regime.

4) The interference threshold constraints have a huge impact on the system performance. For example, when the SNR is 20 dB, the average secrecy sum rate is 0.90 and 0.31 b/s/Hz for $\gamma_1 = 0.2$ and $\gamma_1 = 0.02$, respectively. More specifically, given the set of all feasible precoding matrices, i.e.,

$$\mathcal{P} = \left\{(\mathbf{P}_1, \mathbf{P}_2)\ \middle|\ \begin{array}{l} \mathrm{tr}\left(\mathbf{P}_i^H\mathbf{P}_i\right) \le \beta_i, i = 1, 2 \\ \mathrm{tr}\left(\boldsymbol{\Phi}_{f_j}\right)\cdot\sum\limits_{i=1}^{2}\mathrm{tr}\left(\mathbf{P}_i^H\boldsymbol{\Psi}_{f_{i,j}}\mathbf{P}_i\right) \le \gamma_j\ \forall j \end{array}\right\} \quad (63)$$

we define the following parameters:

$$\bar{\beta}_i = \min\left\{\min_{1\le j\le J}\left\{\frac{\gamma_j}{\mathrm{tr}\left(\boldsymbol{\Phi}_{f_j}\right)\cdot\lambda_{\min}\left(\boldsymbol{\Psi}_{f_{i,j}}\right)}\right\}, \beta_i\right\} \quad (64)$$

where $\lambda_{\min}(\mathbf{A})$ represents the smallest eigenvalue of $\mathbf{A}$. Then, for all $(\mathbf{P}_1, \mathbf{P}_2) \in \mathcal{P}$, we can easily prove that

$$\mathrm{tr}\left(\mathbf{P}_i^H\mathbf{P}_i\right) \le \bar{\beta}_i, \quad i = 1, 2. \quad (65)$$

Equation (65) implies that when $\bar{\beta}_i < \beta_i$, $i = 1, 2$, the power constraints in $\mathcal{P}$ are inactive, i.e., only a portion of the available transmit power can be used to meet all interference threshold constraints. In the case of Figs. 4 and 5, $(\bar{\beta}_1, \bar{\beta}_2)$ is calculated as

$$(\bar{\beta}_1, \bar{\beta}_2) = \begin{cases} (0.5, 0.2), & \gamma_1 = 0.2 \\ (0.05, 0.02), & \gamma_1 = 0.02. \end{cases} \quad (66)$$

Since $(\beta_1, \beta_2) = (2, 2)$, the sum-rate performance in Figs. 4 and 5 is only constrained by interference threshold constraints.

5) The performance of the no-precoding case is very poor because we do not exploit any statistical CSI from STs to the SR and the ED.

C. Comparison of Different Modulations

Finally, we investigate the average secrecy sum rate with different modulations. We consider a secure cognitive radio system with two STs, one SR, one ED, and two PRs. Each node is equipped with two antennas. The correlation matrices are given by

$$\begin{aligned} \boldsymbol{\Phi}_h &= \mathbf{C}(0.3), \boldsymbol{\Psi}_{h_1} = \mathbf{C}(0.9), \boldsymbol{\Psi}_{h_2} = \mathbf{C}(0.95) \\ \boldsymbol{\Phi}_g &= \mathbf{C}(0.6), \boldsymbol{\Psi}_{g_1} = \mathbf{C}(0.7), \boldsymbol{\Psi}_{g_2} = \mathbf{C}(0.2) \\ \boldsymbol{\Phi}_{f_1} &= \mathbf{C}(0.4), \boldsymbol{\Psi}_{f_{1,1}} = \mathbf{C}(0.6), \boldsymbol{\Psi}_{f_{2,1}} = \mathbf{C}(0.4) \\ \boldsymbol{\Phi}_{f_2} &= \mathbf{C}(0.5), \boldsymbol{\Psi}_{f_{1,2}} = \mathbf{C}(0.3), \boldsymbol{\Psi}_{f_{2,2}} = \mathbf{C}(0.5). \end{aligned} \quad (67)$$

The maximum transmission power at the ith ST is given as $\beta_1 = \beta_2 = 2$. The interference threshold at the jth PR is set as $\gamma_1 = \gamma_2 = 0.2$. The noise variance is $\sigma_R^2 = \sigma_E^2 = \sigma^2$.

Fig. 6 shows the average secrecy sum rate with BPSK, QPSK, and 8-PSK modulations. The results in Fig. 6 show that the average secrecy sum rate is an increasing function with respect to the order of modulation. They also indicate that our proposed precoding design can achieve robust performances for a large range of SNRs with different modulations.

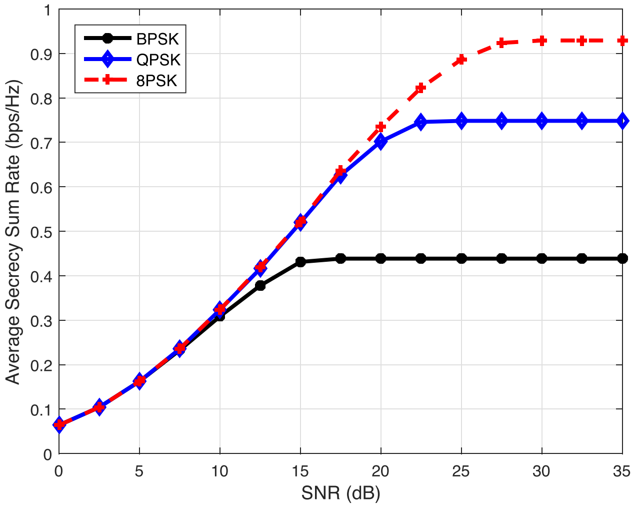


Fig. 6. Average secrecy sum rate for the fading CMAC-WT with different modulations.

V. Conclusion

In this paper, we have considered the precoding design for the fading CMAC-WT with finite-alphabet inputs. We have presented a two-layer precoding algorithm, which exploits statistical CSI of fading channels, to maximize the approximated average secrecy sum rate. The key idea of our algorithm is to find a computationally efficient DC representation of the approximated average secrecy sum rate. By introducing a new matrix variable, we have reformulated the approximated average secrecy sum rate as a DC function and then generated a sequence of relaxed sets to approach the nonconvex feasible set. Each relaxed set can be expressed as the union of convex sets. Finally, near-optimal precoding matrices have been obtained iteratively by maximizing the approximated average secrecy sum rate over a sequence of relaxed sets.

Several numerical results have been provided to demonstrate the efficacy of our proposed precoding algorithm. They have also shown that the proposed precoding algorithm is superior to the conventional Gaussian precoding method and the no-precoding case in the medium- and high-SNR regimes.

Appendix A
Proofs of Propositions 1 and 2

Proof of Proposition 1: We rewrite $\mathcal{F}_{\text{init}}$ as the union of K subsets, i.e.,

$$\mathcal{F}_{\text{init}} = \bigcup_{i=1}^{K} \bar{\mathcal{F}}_i \tag{68}$$

where $\bar{\mathcal{F}}_i$ is given by

$$\bar{\mathcal{F}}_i = \{(\mathbf{Q},\mathbf{p})|(\mathbf{Q},\mathbf{p}) \in \mathcal{F}_{\text{init}}, \mathbf{p} \in \mathcal{B}_i\}. \tag{69}$$

For any $(\mathbf{Q},\mathbf{p}) \in \bar{\mathcal{F}}_i$, the following inequalities hold:

$$(\mathbf{p} - \mathbf{l}(\mathcal{B}_i)) \cdot (\mathbf{p} - \mathbf{l}(\mathcal{B}_i))^T \geq \mathbf{0} \tag{70}$$

$$(\mathbf{p} - \mathbf{u}(\mathcal{B}_i)) \cdot (\mathbf{p} - \mathbf{u}(\mathcal{B}_i))^T \geq \mathbf{0} \tag{71}$$

$$(\mathbf{p} - \mathbf{l}(\mathcal{B}_i)) \cdot (\mathbf{p} - \mathbf{u}(\mathcal{B}_i))^T \leq \mathbf{0} \tag{72}$$

$$\mathbf{Q} = \mathbf{p}\mathbf{p}^T, \mathbf{l}(\mathcal{B}_i) \leq \mathbf{p} \leq \mathbf{u}(\mathcal{B}_i). \tag{73}$$

Thus, $\bar{\mathcal{F}}_i$ can be rewritten as

$$\bar{\mathcal{F}}_i = \{(\mathbf{Q},\mathbf{p})|(\mathbf{Q},\mathbf{p}) \in \mathcal{F}_{\text{init}}, \mathbf{p} \in \mathcal{B}_i\} \cap \mathcal{S}(\mathcal{B}_i). \tag{74}$$

By relaxing $\mathbf{Q} = \mathbf{p}\mathbf{p}^T$ in $\bar{\mathcal{F}}_i$ into $\mathbf{Q} \succeq \mathbf{p}\mathbf{p}^T$, one can easily obtain the following:

$$\bar{\mathcal{F}}_i \subseteq \mathcal{C}(\mathcal{B}_i) \quad \forall i. \tag{75}$$

Therefore, $\mathcal{F}_{\text{init}} \subseteq \mathcal{C}(\mathcal{B}_1) \cup \cdots \cup \mathcal{C}(\mathcal{B}_K)$. This completes the proof. ■

Proof of Proposition 2: We divide $\mathcal{S}(\mathcal{B})$ into two subsets, i.e.,

$$\mathcal{S}(\mathcal{B}) = \mathcal{S}_1(\mathcal{B}) \cup \mathcal{S}_2(\mathcal{B}) \tag{76}$$

where $\mathcal{S}_1(\mathcal{B})$ and $\mathcal{S}_2(\mathcal{B})$ are given by

$$\begin{aligned} \mathcal{S}_1(\mathcal{B}) &= \{(\mathbf{Q},\mathbf{p})|(\mathbf{Q},\mathbf{p}) \in \mathcal{S}(\mathcal{B}), \mathbf{p} \in \mathcal{B}_1\} \\ \mathcal{S}_2(\mathcal{B}) &= \{(\mathbf{Q},\mathbf{p})|(\mathbf{Q},\mathbf{p}) \in \mathcal{S}(\mathcal{B}), \mathbf{p} \in \mathcal{B}_2\}. \end{aligned} \tag{77}$$

It is obvious that if we can prove

$$\begin{aligned} \mathcal{S}(\mathcal{B}_1) &\subseteq \mathcal{S}_1(\mathcal{B}) \\ \mathcal{S}(\mathcal{B}_2) &\subseteq \mathcal{S}_2(\mathcal{B}) \end{aligned} \tag{78}$$

then $\mathcal{C}(\mathcal{B}_1) \cup \mathcal{C}(\mathcal{B}_2) \subseteq \mathcal{C}(\mathcal{B})$. We will restrict our attention to show $\mathcal{S}(\mathcal{B}_1) \subseteq \mathcal{S}_1(\mathcal{B})$, and $\mathcal{S}(\mathcal{B}_2) \subseteq \mathcal{S}_2(\mathcal{B})$ can be proved in the same way.

Since $\mathcal{B}_1 \subseteq \mathcal{B}$, we have

$$\mathbf{l}(\mathcal{B}) \leq \mathbf{l}(\mathcal{B}_1) \leq \mathbf{u}(\mathcal{B}_1) \leq \mathbf{u}(\mathcal{B}). \tag{79}$$

Therefore, the following inequalities hold for any $\mathbf{l}(\mathcal{B}_1) \leq \mathbf{p} \leq \mathbf{u}(\mathcal{B}_1)$:

$$[\mathbf{l}(\mathcal{B}_1) - \mathbf{l}(\mathcal{B})][\mathbf{p} - \mathbf{l}(\mathcal{B}_1)]^T + [\mathbf{p} - \mathbf{l}(\mathcal{B})]\,[\mathbf{l}(\mathcal{B}_1) - \mathbf{l}(\mathcal{B})]^T \geq \mathbf{0}$$

$$[\mathbf{u}(\mathcal{B}_1) - \mathbf{u}(\mathcal{B})][\mathbf{p} - \mathbf{u}(\mathcal{B}_1)]^T + [\mathbf{p} - \mathbf{u}(\mathcal{B})][\mathbf{u}(\mathcal{B}_1) - \mathbf{u}(\mathcal{B})]^T \geq \mathbf{0}$$

$$[\mathbf{l}(\mathcal{B}_1) - \mathbf{l}(\mathcal{B})][\mathbf{p} - \mathbf{u}(\mathcal{B})]^T + [\mathbf{p} - \mathbf{l}(\mathcal{B}_1)][\mathbf{u}(\mathcal{B}_1) - \mathbf{u}(\mathcal{B})]^T \leq \mathbf{0}.$$

The given inequalities can be rewritten, respectively, as

$$\begin{aligned} &\mathbf{Q} - \mathbf{L}_{\mathbf{p}}(\mathcal{B}) - \mathbf{L}_{\mathbf{p}}(\mathcal{B})^T + \mathbf{l}(\mathcal{B}) \cdot \mathbf{l}(\mathcal{B})^T \\ &\quad \geq \mathbf{Q} - \mathbf{L}_{\mathbf{p}}(\mathcal{B}_1) - \mathbf{L}_{\mathbf{p}}(\mathcal{B}_1)^T + \mathbf{l}(\mathcal{B}_1) \cdot \mathbf{l}(\mathcal{B}_1)^T \\ &\mathbf{Q} - \mathbf{U}_{\mathbf{p}}(\mathcal{B}) - \mathbf{U}_{\mathbf{p}}(\mathcal{B})^T + \mathbf{u}(\mathcal{B}) \cdot \mathbf{u}(\mathcal{B})^T \\ &\quad \geq \mathbf{Q} - \mathbf{U}_{\mathbf{p}}(\mathcal{B}_1) - \mathbf{U}_{\mathbf{p}}(\mathcal{B}_1)^T + \mathbf{u}(\mathcal{B}_1) \cdot \mathbf{u}(\mathcal{B}_1)^T \\ &\mathbf{Q} - \mathbf{L}_{\mathbf{p}}(\mathcal{B}) - \mathbf{U}_{\mathbf{p}}(\mathcal{B})^T + \mathbf{l}(\mathcal{B}) \cdot \mathbf{u}(\mathcal{B})^T \\ &\quad \leq \mathbf{Q} - \mathbf{L}_{\mathbf{p}}(\mathcal{B}_1) - \mathbf{U}_{\mathbf{p}}(\mathcal{B}_1)^T + \mathbf{l}(\mathcal{B}_1) \cdot \mathbf{u}(\mathcal{B}_1)^T \end{aligned} \tag{80}$$

where $\mathbf{L}_{\mathbf{p}}(\mathcal{B}) = \mathbf{l}(\mathcal{B}) \cdot \mathbf{p}^T$, and $\mathbf{U}_{\mathbf{p}}(\mathcal{B}) = \mathbf{u}(\mathcal{B}) \cdot \mathbf{p}^T$. Inequalities (80) provide a sufficient condition for $\mathcal{S}(\mathcal{B}_1) \subseteq \mathcal{S}_1(\mathcal{B})$. Therefore, $\mathcal{C}(\mathcal{B}_1) \cup \mathcal{C}(\mathcal{B}_2) \subseteq \mathcal{C}(\mathcal{B})$. This completes the proof. ■

APPENDIX B
PROOF OF PROPOSITION 3

Since $\{\varphi(\mathcal{F}_k)\}$ is a monotonically decreasing sequence lower bounded by $\varphi(\mathcal{F}_{\text{init}})$, the limit of $\{\varphi(\mathcal{F}_k)\}$ exists [36]. Suppose that

$$\lim_{k\to\infty} \varphi(\mathcal{F}_k) = v > \varphi(\mathcal{F}_{\text{init}}) \tag{81}$$

then for any $\varepsilon > 0$, there exists $K > 0$ such that for any $k > K$, we have

$$v < \varphi\left(\mathcal{C}(\mathcal{B}_g)\right) < v + \varepsilon. \tag{82}$$

Let $r(\mathcal{B})$ denote the length of the longest edge of a hyperrectangle $\mathcal{B}$ satisfying $\mathcal{B} \subseteq \mathcal{B}_{\text{init}}$. In each iteration of Algorithm 1, we divide $\mathcal{B}_g$ along $r(\mathcal{B}_g)$ into two hyperrectangles. Therefore, $r(\mathcal{B}_g)$ should satisfy the following condition:

$$\lim_{k\to\infty} r(\mathcal{B}_g) = 0. \tag{83}$$

We further denote the center of $\mathcal{B}_g$ by $\mathbf{p}_g$, i.e., $\mathbf{p}_g = (\mathbf{l}(\mathcal{B}_g) + \mathbf{u}(\mathcal{B}_g))/2$. When $r(\mathcal{B}_g) \to 0$, we have

$$\mathcal{S}(\mathcal{B}_g) \to \left\{(\mathbf{Q}, \mathbf{p}) \,\middle|\, \mathbf{Q} = \mathbf{p}_g \mathbf{p}_g^T, \mathbf{p} = \mathbf{p}_g \right\}. \tag{84}$$

Therefore, $\mathcal{C}(\mathcal{B}_g)$ converges to a point when $\mathbf{p}_g$ belongs to the feasible set $\mathcal{P}$; otherwise, $\mathcal{C}(\mathcal{B}_g)$ is an empty set. Thus, we have

$$\lim_{r(\mathcal{B}_g)\to 0} \varphi\left(\mathcal{C}(\mathcal{B}_g)\right) = f(\mathbf{p}_g) - g(\mathbf{p}_g), \quad \mathbf{p}_g \in \mathcal{P}. \tag{85}$$

Combining (83) and (85), we conclude that

$$\lim_{k\to\infty} \varphi\left(\mathcal{C}(\mathcal{B}_g)\right) = f(\mathbf{p}_g) - g(\mathbf{p}_g) < v \tag{86}$$

which is contradictory to (82). Therefore, $\{\varphi(\mathcal{F}_k)\}$ converges to $\varphi(\mathcal{F}_{\text{init}})$. This completes the proof. ■

REFERENCES

[1] A. Goldsmith, S. A. Jafar, I. Marić, and S. Srinivasa, “Breaking spectrum gridlock with cognitive radios: An information theoretic perspective,” *Proc. IEEE*, vol. 97, no. 5, pp. 894–914, May 2009.

[2] L. Zhang, Y.-C. Liang, and Y. Xin, “Joint beamforming and power allocation for multiple access channels in cognitive radio networks,” *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 38–51, Jan. 2008.

[3] L. Zhang, Y. Xin, and Y.-C. Liang, “Weighted sum rate optimization for cognitive radio MIMO broadcast channels,” *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 2950–2959, Jun. 2009.

[4] H. Chan, A. Perrig, and D. Song, “Random key predistribution schemes for sensor networks,” in *Proc. IEEE Symp. SP*, 2003, pp. 197–213.

[5] C. E. Shannon, “Communication theory of secrecy systems,” *Bell Syst. Tech. J.*, vol. 28, no. 4, pp. 656–715, Oct. 1949.

[6] A. D. Wyner, “The wire-tap channel,” *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, Oct. 1975.

[7] I. Csiszár and J. Korner, “Broadcast channels with confidential messages,” *IEEE Trans. Inf. Theory*, vol. IT-24, no. 3, pp. 339–348, May 1978.

[8] E. Tekin and A. Yener, “The Gaussian multiple access wire-tap channel,” *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5747–5755, Dec. 2008.

[9] E. Tekin and A. Yener, “The general Gaussian multiple-access and two-way wiretap channels: Achievable rates and cooperative jamming,” *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2735–2751, Jun. 2008.

[10] G. Geraci, M. Egan, J. Yuan, A. Razi, and I. B. Collings, “Secrecy sum-rates for multi-user MIMO regularized channel inversion precoding,” *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3472–3482, Nov. 2012.

[11] G. Geraci, R. Couillet, J. Yuan, M. Debbah, and I. B. Collings, “Large system analysis of linear precoding in MISO broadcast channels with confidential messages,” *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 1660–1671, Sep. 2013.

[12] N. Yang, G. Geraci, J. Yuan, and R. Malaney, “Confidential broadcasting via linear precoding in non-homogeneous MIMO multiuser networks,” *IEEE Trans. Commun.*, vol. 62, no. 7, pp. 2515–2530, Jul. 2014.

[13] J. Yang, I.-M. Kim, and D. I. Kim, “Joint design of optimal cooperative jamming and power allocation for linear precoding,” *IEEE Trans. Commun.*, vol. 62, no. 9, pp. 3285–3298, Sep. 2014.

[14] M. F. Hanif, L.-N. Tran, M. Juntti, and S. Glisic, “On linear precoding strategies for secrecy rate maximization in multiuser multiantenna wireless networks,” *IEEE Trans. Signal Process.*, vol. 62, no. 14, pp. 3536–3551, Jul. 2014.

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

[16] C. Xiao and Y. R. Zheng, “On the mutual information and power allocation for vector Gaussian channels with finite discrete inputs,” in *Proc. IEEE GLOBECOM*, 2008, pp. 1–5.

[17] F. Pérez-Cruz, M. R. 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.

[18] 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.

[19] M. Wang, W. Zeng, and C. Xiao, “Linear precoding for MIMO multiple access channels with finite discrete inputs,” *IEEE Trans. Wireless Commun.*, vol. 10, no. 11, pp. 3934–3942, Nov. 2011.

[20] W. Zeng, C. Xiao, J. Lu, and K. B. Letaief, “Globally optimal precoder design with finite-alphabet inputs for cognitive radio networks,” *IEEE J. Sel. Areas Commun.*, vol. 30, no. 10, pp. 1861–1874, Nov. 2012.

[21] W. Zeng, C. Xiao, M. Wang, and J. Lu, “Linear precoding for finite-alphabet inputs over MIMO fading channels with statistical CSI,” *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 3134–3148, Jun. 2012.

[22] S. Bashar, Z. Ding, and C. Xiao, “On secrecy rate analysis of MIMO wiretap channels driven by finite-alphabet input,” *IEEE Trans. Commun.*, vol. 60, no. 12, pp. 3816–3825, Dec. 2012.

[23] Y. Wu, C. Xiao, Z. Ding, X. Gao, and S. Jin, “Linear precoding for finite-alphabet signaling over MIMOME wiretap channels,” *IEEE Trans. Veh. Technol.*, vol. 61, no. 6, pp. 2599–2612, Jul. 2012.

[24] J. Harshan and B. S. Rajan, “A novel power allocation scheme for two-user GMAC with finite input constellations,” *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 818–827, Feb. 2013.

[25] S. Vishwakarma and A. Chockalingam, “Decode-and-forward relay beamforming for secrecy with finite-alphabet input,” *IEEE Commun. Lett.*, vol. 17, no. 5, pp. 912–915, May 2013.

[26] S. Vishwakarma and A. Chockalingam, “Power allocation in MIMO wiretap channel with statistical CSI and finite-alphabet input,” in *Proc. Nat. Conf. Commun.*, 2014, pp. 1–6.

[27] M. Girnyk, M. Vehkapera, and L. K. Rasmussen, “Large-system analysis of correlated MIMO multiple access channels with arbitrary signaling in the presence of interference,” *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 2060–2073, Apr. 2014.

[28] W. Zeng, Y. Zheng, and C. Xiao, “Multi-antenna secure cognitive radio networks with finite-alphabet inputs: A global optimization approach for precoder design,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 3044–3057, Apr. 2016.

[29] R. Horst and H. Tuy, *Global Optimization: Deterministic Approaches*. Berlin, Germany: Springer-Verlag, 1996.

[30] P. D. Tao *et al.*, “Duality in DC (difference of convex functions) optimization. Subgradient methods,” in *Trends in Mathematical Optimization*. Berlin, Germany: Springer-Verlag, 1988, pp. 277–293.

[31] A. L. Yuille and A. Rangarajan, “The concave–convex procedure,” *Neural Comput.*, vol. 15, no. 4, pp. 915–936, 2003.

[32] A. Ferrer and J. E. Martínez-Legaz, “Improving the efficiency of DC global optimization methods by improving the DC representation of the objective function,” *J. Global Optim.*, vol. 43, no. 4, pp. 513–531, Aug. 2009.

[33] C. Xiao, J. Wu, S.-Y. 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.

[34] W. Xu, X. Dong, and W.-S. Lu, “Joint precoding optimization for multiuser multi-antenna relaying downlinks using quadratic programming,” *IEEE Trans. Commun.*, vol. 59, no. 5, pp. 1228–1235, May 2011.

[35] G. A. Seber, *A Matrix Handbook for Statisticians*, vol. 15. New York, NY, USA: Wiley, 2008.

[36] W. Rudin, *Principles of Mathematical Analysis*, vol. 3. New York, NY, USA: McGraw-Hill, 1964.

[37] A. Nemirovski, C. Roos, and T. Terlaky, “On maximization of quadratic form over intersection of ellipsoids with common center,” *Math. Program.*, vol. 86, no. 3, pp. 463–473, 1999.

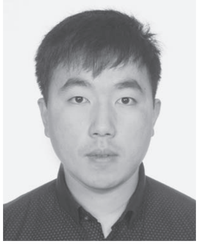
[38] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[39] A. Beck, A. Ben-Tal, and L. Tetruashvili, "A sequential parametric convex approximation method with applications to nonconvex truss topology design problems," *J. Global Optim.*, vol. 47, no. 1, pp. 29–51, 2010.

[40] O. Mehanna, K. Huang, B. Gopalakrishnan, A. Konar, and N. Sidiropoulos, "Feasible point pursuit and successive approximation of non-convex QCQPs," *IEEE Signal Process. Lett.*, vol. 22, no. 7, pp. 804–808, Jul. 2015.

[41] A. Khabbazibasmenj, F. Roemer, S. A. Vorobyov, and M. Haardt, "Sum-rate maximization in two-way AF MIMO relaying: Polynomial time solutions to a class of DC programming problems," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5478–5493, Oct. 2012.

[42] Z.-Q. Luo, W.-K. Ma, A.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.



Juening Jin received the B.S. degree in electronic engineering from Southeast University, Nanjing, China, in 2013. He is currently working toward the Ph.D. degree with the Network Coding and Transmission Laboratory, Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China.

He is currently a Visiting Scholar with Missouri University of Science and Technology, Rolla, MO, USA. His research interests include convex and non-convex optimizations and physical-layer security.

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.

He is a Professor with the Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO, USA. He is currently a Program Director with the National Science Foundation through intergovernmental personnel act assignment. He was a Senior Member of Scientific Staff with Nortel Networks, Ottawa, ON, Canada, and a faculty member with Tsinghua University; the University of Alberta, Edmonton, AB, Canada; and the University of Missouri, Columbia, MO. He has also held visiting professor positions in Germany and Hong Kong. He is the holder of three U.S. patents. His invented algorithms have been implemented into Nortel's base station radio products after successful technical field trials and network integration. His research interests include wireless communications, signal processing, and underwater acoustic communications.

Dr. Xiao served as an elected member of the Board of Governors, a member of the Fellow Evaluation Committee, the Director of Conference Publications, and a Distinguished Lecturer for the IEEE Communications Society. He also served as an Editor, an Area Editor, and the Editor-in-Chief for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and as an Associate Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—I. He was the Technical Program Chair of the 2010 IEEE International Conference on Communications in Cape Town, South Africa. He served as the founding Chair of the IEEE Wireless Communications Technical Committee. He has received several distinguished awards, including the 2014 Humboldt Research Award, the 2014 IEEE Communications Society Joseph LoCicero Award, and the 2015 IEEE Wireless Communications Technical Committee Recognition Award.

Meixia Tao (SM'10) received the B.S. degree in electronic engineering from Fudan University, Shanghai, China, in 1999 and the Ph.D. degree in electrical and electronic engineering from the Hong Kong University of Science and Technology, Hong Kong, in 2003.

She is currently a Professor with the Department of Electronic Engineering, Shanghai Jiao Tong University. During 2003–2004, she was a Member of Professional Staff with the Hong Kong Applied Science and Technology Research Institute. From 2004 to 2007, she was a Teaching Fellow and then an Assistant Professor with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore. Her current research interests include content-centric wireless networks, resource allocation, interference management, and physical-layer security.

Dr. Tao currently serves as a member of the Executive Editorial Committee of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS. She was on the Editorial Board of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2007–2010), the IEEE COMMUNICATIONS LETTERS (2009–2012), and the IEEE WIRELESS COMMUNICATIONS LETTERS (2011–2015). She also served as a Technical Program Committee Chair for the 2014 IEEE/CIC International Conference on Communications in China (ICCC) and a Symposium Cochair for the 2015 IEEE International Conference on Communications. She received the IEEE Heinrich Hertz Award for Best Communications Letters in 2013, the IEEE/CIC ICCC Best Paper Award in 2015, and the International Conference on Wireless Communications and Signal Processing Best Paper Award in 2012. She also received the IEEE Communications Society Asia-Pacific Outstanding Young Researcher Award in 2009.

Wen Chen (SM'11) received the B.S. and M.S. degrees from Wuhan University, Wuhan, China, in 1990 and 1993, respectively, and the Ph.D. degree from the University of Electro-Communications, Tokyo, Japan, in 1999.

From 1999 to 2001, he was a Researcher with the Japan Society for the Promotion of Science. In 2001, he joined the University of Alberta, Edmonton, AB, Canada, starting as a Postdoctoral Fellow with the Information Research Laboratory and continuing as a Research Associate with the Department of Electrical and Computer Engineering. Since 2006, he has been a Full Professor with the Department of Electronic Engineering, Shanghai Jiao Tong University (SJTU), Shanghai, China, where he is also the Director of the Institute for Signal Processing and Systems. During 2014–2015, he was the Dean of Electronic Engineering and Automations with Guilin University of Electronic Technology, Guilin, China. Since 2016, he has been the Chairman of SJTU Intelligent Property Management Corporation. He is the author of 73 papers in IEEE JOURNALS AND TRANSACTIONS and more than 100 papers in IEEE conference publications. His research interests include network coding, cooperative communications, cognitive radio, and multiple-input–multiple-output orthogonal frequency-division multiplexing systems.