

Secure Communications for Dual-Polarized MIMO Systems

Shiqi Gong, Chengwen Xing, Sheng Chen, *Fellow, IEEE*, and Zesong Fei

Abstract—To enhance secure communications, we deploy the dual-polarized antenna arrays at communication nodes of the multi-input multi-output (MIMO) system where the base station communicates with multiple legitimate users in the presence of an eavesdropper. We also adopt the dual-structured precoding in which a preprocessing matrix based on the polarized array spatial correlation and a linear precoding based on the instantaneous channel state information (CSI) are concatenated. We design this dual-structured multi-user linear precoding under three cases. In the first case, given perfect global CSI, the secrecy rate optimization problem is formulated and transformed into the weighted minimum mean square error problem, which can be effectively solved by the block coordinate decent method. In the second case where the eavesdropper's CSI is unavailable, an artificial noise is generated to confuse the eavesdropper by minimizing the information transmit power subject to a preset mean square error (MSE) threshold for the recovered confidential signals, which can be solved by an efficient iterative algorithm. In the third case of imperfect global CSI, the robust optimization for secure communications is performed by minimizing the largest received MSE among the users subject to the total transmit power constraint, which can be reformulated into a biconvex semidefinite programming problem and solved by an efficient alternating convex optimization. Simulation results are included to demonstrate the excellent performance of our proposed designs over the conventional single-polarized array based designs, in terms of achievable secrecy rate, minimum transmit power and the MSE of recovered confidential signals.

Index Terms—Physical layer security, dual-polarized MIMO system, dual-structured multi-user linear precoding

I. INTRODUCTION

Multiple-input multiple-output (MIMO) offers a key technology to support higher data rate and to improve the energy and spectral efficiency. In order to attain the available maximum multiplexing and/or diversity gains in MIMO systems, the antenna spacing must be at least ten wavelengths [1]–[3]. Consequently, the deployment of antenna arrays, particularly large-scale arrays, may be difficult due to the device space limitation. Even when an antenna array can be deployed, the unavoidable mutual coupling and high spatial correlation among antenna elements are not negligible [4]. The multi-polarized antenna array is proposed as an effective solution, which exploits the polarization difference of the electromagnetic (EM) waves in wireless channels to reduce the array spatial correlation and to realize the robust and compact communication devices [5], [6]. In particular, it was demonstrated that the dual-polarized antenna elements have low spatial correlation, compared to the conventional single-polarized antenna array, owing to the fact that the orthogonally

polarized EM waves experience independent fading in both line-of-sight (LOS) and non-line-of-sight (NLOS) scenarios [7], [8]. Moreover, given the same number of antenna elements and with the same spacing among co-polarized array elements, the size of a single-polarized antenna array is twice that of a dual-polarized antenna array. Hence, in our work, the dual-polarized array is deployed at each communication node of the downlink (DL) MIMO system.

When the MIMO system operates in the frequency division duplexing (FDD) mode, where the uplink and DL channel reciprocity does not hold, the DL channel acquisition consumes huge amount of training overhead [9]. For conventional single-polarized MIMO, to reduce the burden of DL channel acquisition at base station (BS), a dual-structured linear precoding scheme was proposed, in which a preprocessing operation based on the spatial covariance matrix and a linear precoding utilizing the instantaneous channel state information (CSI) are concatenated [10], [11]. Generally, the spatial correlation varies slowly, compared to the instantaneous CSI, and thus it can be obtained accurately with a low feedback cost [11]. Therefore, we also adopts the dual-structured multi-user linear precoding as an effective means of reducing the channel feedback overhead for dual-polarized MIMO systems. As a benefit of the orthogonally polarized channel characteristics, the spatial covariance matrix of the dual-polarized MIMO channel is block-diagonal and, consequently, the design of the dual-structured precoding for the dual-polarized MIMO has a lower complexity than that for the single-polarized MIMO.

Traditionally, the multi-user linear precoding scheme has been widely adopted in MIMO systems for effective information transmission. For example, in [12] a linear precoding scheme was proposed to improve the multi-user MIMO system's error performance by relaxing the signal-to-leakage-and-noise (SLNR) maximization. The work [13] jointly designed the optimal linear precoder and decoder for MIMO channels based on a weighted minimum mean square error (WMMSE) criterion under the transmit power constraint. In [14], a simple matched-filter precoding was applied to enable large power efficiency gain for massive MIMO. However, for multi-polarized antenna based MIMO systems, less attention has been paid to the multi-user linear precoding design and, therefore, the prominent advantage of polarized array may be under-utilized. In [15], a dual-structured linear precoding combining with the subgrouping method based on polarization was proposed for the dual-polarized MIMO system to achieve the maximum system capacity. The work [16] designed an adaptive channel subspace sampling algorithm to estimate the millimeter wave beam-alignment vectors for the multi-users in the dual-polarized MIMO system. Our main objective is however focusing on physical layer security.

Owing to the openness of wireless links, any receiver located within the communication range of the transmitter can receive the transmitted signal naturally. As a result, the security

S. Gong, C. Xing (corresponding author) and Z. Fei are with School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (E-mails: gsqyx@163.com, xingchengwen@gmail.com, feizesong@bit.edu.cn).

S. Chen is with School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (E-mail: sqc@ecs.soton.ac.uk), and also with King Abdulaziz University, Jeddah 21589, Saudi Arabia

issue is a paramount concern in wireless communications [17], [18]. The physical layer security has emerged recently as an effective means of avoiding malicious wiretap as far as possible. The concept of physical layer security was firstly defined by Shannon [19]. Then Wyner [20] introduced the wiretap channel model and defined the secrecy capacity at which the information transmission is reliable irrespective of the amount of information leakage to the eavesdropper. Expanding the work of [20], a Gaussian degraded wiretap channel was investigated in [21]. Based on these theoretical fundamentals, a large amount of literature were produced focusing on the physical layer security of various multiple-antenna systems [22]–[28]. Specifically, a security beamforming strategy was designed for the multiple-input single-output (MISO) system in [22], [23], the single-input multiple-output (SIMO) system in [24], and the MIMO system in [25]–[28]. In particular, to effectively balance the security performance among different links of an MIMO network, the work [25] proposed a game theory based collaborative transmission scheme, while an artificial noise scheme was proposed in [26] for secure communications of the two-way relaying MIMO network under a realistic scenario that the CSI of eavesdropper is unavailable. Further extending these research, the work [27] jointly designed a matched-filter data precoding and a null-space or random artificial noise scheme for secure communications of the multi-cell massive MIMO system, while in [28], several data precoding schemes were considered to investigate the achievable ergodic secrecy capacity region of massive MIMO.

There also exist some literature considering the security issue in the multi-polarized system [29], [30]. For example, the work [29] introduced a simple polarization encoding scheme to ensure that the spatially separated multi-receivers can operate in parallel secretly. To the best knowledge of the authors, however, there has been no related work addressing the physical-layer security of the dual-polarized DL MIMO system so far. Hence, it is significant and valuable to consider the dual-structured multi-user linear precoding design for secure communications in the dual-polarized DL MIMO system, which motivates our current work.

In this paper, we model the dual-polarized DL MIMO, in which a BS with a dual-polarized antenna array transmits multiple data streams to multiple mobile users (MUs), each equipped with a dual-polarized array, at the presence of an eavesdropper that is also equipped with a dual-polarized array. We adopt a dual-structured multi-user linear precoding scheme, which includes the preprocessing based on the spatial covariance matrix and the subsequent linear precoding based on the instantaneous CSI, to realize secure communications of the dual-polarized DL MIMO system. The spatial correlation based preprocessing matrix is optimized by utilizing the block diagonalization, while the design of the linear precoding depends on whether the eavesdropper's CSI is available, which is discussed in details in the following three cases.

1) Under the ideal scenario that the perfect global-system CSI is available, the secrecy rate maximization problem is formulated, which is however nonconvex and difficult to solve directly. In order to solve this challenging optimization problem, the WMMSE based equivalent transformation is

conducted and a block coordinate decent (BCD) iterative algorithm is proposed to achieve the maximum secrecy rate.

2) Under the realistic scenario where the CSI of eavesdropper is unavailable, an artificial noise scheme is applied to interfere with the eavesdropper for the sake of achieving secure communications. Because the artificial noise is constrained by the total transmit power, in designing this scheme, we minimize the information transmit power subject to a preset mean square error (MSE) threshold for the recovered confidential signals. An iterative algorithm is proposed to decompose this optimization problem into a standard quadratically constrained quadratic programming (QCQP) subproblem and a closed-form minimum MSE (MMSE) based receiver filter design.

3) Further considering practical situations with imperfect global system CSI which arise owing to channel estimation or feedback error, the worst-case optimization is performed to achieve secure communications, in which the largest received MSE among the MUs is minimized subject to the total transmit power constraint. By reformulating this challenging optimization problem into a biconvex semidefinite programming (SDP) problem, an alternating convex optimization (ACO) algorithm is proposed to perform the SDP optimization separately for different groups of optimization variables.

Our simulation results show that all the aforementioned iterative algorithms, corresponding to different availabilities of CSI, achieve good convergence performance. We also explicitly compare the achievable security performance of the proposed dual-structured precoding for dual-polarized MIMO with those of the dual-structured precoding for conventional single-polarized MIMO under all the three CSI scenarios. Our simulation results confirm that the former outperforms the latter, in terms of the secrecy rate, the minimum transmit power and the MSE of recovered confidential signals. Thus, in addition to providing more compact antenna array which is beneficial for practical implementation as well as imposing lower computational complexity in designing the dual-structured precoding, the dual-polarized MIMO offers the performance advantage, compared to the single-polarized MIMO. Additionally, we apply the standard MIMO precoding schemes [26], [28], [31] to the dual-polarized MIMO system and use their achievable security performance as the benchmarks for the three respective CSI scenarios. Although outperforming the proposed dual-structured precoding scheme, these standard MIMO designs require the full instantaneous CSI which is difficult to acquire in practice and they impose huge channel feedback overhead. Moreover, these standard MIMO precoding schemes impose much higher computational complexity than the proposed dual-structured precoding design.

The rest of this paper is organized as follows. Section II introduces the dual-polarized system model and the dual-structured multi-user linear precoding. In Section III, the dual-structured multi-user linear precoding designs given the perfect CSI of all BS-MU links are presented for the ideal scenario of perfectly available eavesdropper's CSI and the realistic case where the eavesdropper's CSI is unavailable. Section IV details our proposed robust design for achieving secure communications under the practical scenario of imperfect global CSI. The simulation results and our conclusions are given in Sections V

and VI, respectively.

We adopt the following notation conventions. The normal-faced lower-case letters denote scalars, while bold-faced lower-case and upper-case letters denote vectors and matrices, respectively. The imaginary axis is given by $j = \sqrt{-1}$. The transpose, conjugate transpose, inverse and pseudo-inverse operators are denoted by $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^{-1}$ and $(\cdot)^\dagger$, respectively, while $|\cdot|$, $\|\cdot\|$ and $\|\cdot\|_F$ denote the absolute value, Euclidean and Frobenius norms, respectively. The operator \otimes denotes the Kronecker product and \odot denotes the elementwise multiplication of two matrices, while $\text{Tr}(\cdot)$ and $\det(\cdot)$ denote the matrix trace and determinant operators, respectively. \mathbf{I}_N is the $N \times N$ identity matrix, and $\mathbf{0}$ is the zero matrix/vector of appropriate dimension, while $\mathbf{1}_{m \times n}$ denotes the $m \times n$ matrix with every element equaling to 1. $\mathbf{A} \succ 0$ and $\mathbf{A} \succeq 0$ indicate that \mathbf{A} is positive and semipositive definite matrix, respectively, while $\mathbf{A} \succeq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is semipositive definite. $E[\cdot]$ denotes the expectation operator. $[\mathbf{A}]_{m:n}$ is the sub-matrix containing all elements from the m th column to the n th column of \mathbf{A} , and $\text{diag}\{a_1, \dots, a_m\}$ is the diagonal matrix with the diagonal elements a_1, \dots, a_m , while $\text{Bdiag}\{\mathbf{A}_1, \dots, \mathbf{A}_m\}$ is the block diagonal matrix with $\mathbf{A}_1, \dots, \mathbf{A}_m$ at its block diagonal positions. The i th-row and j th-column element of \mathbf{A} is given by $A_{i,j}$. \mathbf{H}_\perp denotes the orthogonal projection matrix onto the null space of \mathbf{H} , and $[a]^+ = \max\{a, 0\}$. Table I lists the main variables used in this work.

TABLE I
LIST OF MAIN VARIABLES.

Variable	Definition
N_t, N_r, N_e	Numbers of antennas at BS, MUs, eavesdropper
K	Number of MUs
\mathbf{s}_k	Desired data streams for k th MU
$\mathbf{H}_k, \mathbf{H}_e$	Channel matrices from BS to k th MU, eavesdropper
\mathbf{V}_k	Precoding matrix for k th MU
\mathbf{B}_k	Preprocessing matrix in dual-structured precoding
\mathbf{P}_k	Linear precoding matrix in dual-structured precoding
$\mathbf{H}_{k,j}$	Virtual channel j th MU \rightarrow k th MU in perfect CSI
$\mathbf{W}_k, \mathbf{N}_k$	Auxiliary variables of the BCD method
$\mathbf{H}_{k,j}$	Virtual channel j th MU \rightarrow k th MU in unknown CSI
\mathbf{M}_k	Receive filter for k th MU
\mathbf{MSE}_k	Covariance matrix corresponding to \mathbf{M}_k
$\widehat{\mathbf{H}}_k, \widehat{\mathbf{H}}_e$	Nominal channels BS \rightarrow k th MU and eavesdropper
Δ_k, Δ_e	Channel uncertainties for k th MU and eavesdropper

II. SYSTEM MODEL AND DUAL-STRUCTURED LINEAR PRECODING

Without loss of generality, we consider a single-cell DL system which consists of one BS equipped with N_t polarized antennas, K legitimate MUs, each having N_r polarized antennas, and one eavesdropper with N_e polarized antennas. Here, N_t, N_r and N_e are all even numbers. As shown in Fig. 1, the BS, each MU and eavesdropper are deployed with $N_t/2, N_r/2$ and $N_e/2$ pairs of collocated vertically/horizontally polarized antennas, respectively. Based on this structure, the received signal vector $\mathbf{y}_k \in \mathbb{C}^{N_r}$ at the k th MU is expressed as

$$\mathbf{y}_k = \begin{bmatrix} \mathbf{y}_k^v \\ \mathbf{y}_k^h \end{bmatrix} = \mathbf{H}_k^H \mathbf{x} + \mathbf{n}_k, \quad (1)$$

where $\mathbf{y}_k^v \in \mathbb{C}^{N_r/2}$ and $\mathbf{y}_k^h \in \mathbb{C}^{N_r/2}$ denote the received signals at the k th MU with the vertical and horizontal polarizations, respectively, $\mathbf{n}_k \in \mathbb{C}^{N_r}$ represents the channel additive

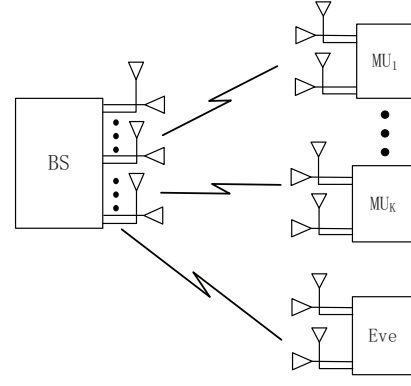


Fig. 1. The dual-polarized MIMO downlink system.

white Gaussian noise (AWGN) vector with $E[\mathbf{n}_k \mathbf{n}_k^H] = \sigma_k^2 \mathbf{I}_{N_r}$, i.e., $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \sigma_k^2 \mathbf{I}_{N_r})$, and $\mathbf{H}_k \in \mathbb{C}^{N_t \times N_r}$ denotes the channel matrix from the BS to the k th MU, while $\mathbf{x} \in \mathbb{C}^{N_t}$ is the linearly precoded transmit signal vector for the K MUs. Given the data vector $\mathbf{s}_k \in \mathbb{C}^{d_k}$ and the precoding matrix $\mathbf{V}_k \in \mathbb{C}^{N_t \times d_k}$ for the k th MU, where $1 \leq k \leq K$, \mathbf{x} is generally expressed as

$$\mathbf{x} = \sum_{k=1}^K \mathbf{V}_k \mathbf{s}_k. \quad (2)$$

For simplicity, we assume $d_1 = \dots = d_K = d_s$. Furthermore, the maximum transmit power P_{\max} is imposed on \mathbf{x} as $E[\|\mathbf{x}\|^2] = \sum_{k=1}^K \text{Tr}(\mathbf{V}_k \mathbf{V}_k^H) \leq P_{\max}$.

A. Polarized channel modeling

Similar to [32], we assume that the spatial covariance matrices are the same for the both polarizations.¹ For the k th MU, denote its polarized spatial covariance matrix² by $\mathbf{R}_k^p \in \mathbb{C}^{\frac{N_r}{2} \times \frac{N_r}{2}}$. Further arranging the r_k non-zero eigenvalues of \mathbf{R}_k^p in decreasing order of magnitude $\gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_{r_k}$, where generally $r_k < \frac{N_r}{2}$, we have the diagonal matrix $\Lambda_k = \text{diag}\{\gamma_1, \dots, \gamma_{r_k}\}$ and the unitary matrix $\mathbf{U}_k \in \mathbb{C}^{\frac{N_r}{2} \times r_k}$ which consists of the corresponding eigenvectors of \mathbf{R}_k^p .

According to [6], [15], the dual-polarized DL MIMO channel matrix \mathbf{H}_k can be expressed as

$$\mathbf{H}_k = \left(\begin{bmatrix} 1 & r_p \\ r_p & 1 \end{bmatrix} \otimes (\mathbf{U}_k \Lambda_k^{\frac{1}{2}}) \right) \left(\mathbf{G}_k \odot (\mathbf{X} \otimes \mathbf{1}_{r_k \times \frac{N_t}{2}}) \right), \quad (3)$$

where r_p is the correlation coefficient between the horizontally and vertically polarized antennas which is generally closed to zero ($r_p \approx 0$) according to [6], and \mathbf{G}_k , whose elements follow the circularly symmetric Gaussian distribution with zero mean and unit power, are defined by

$$\mathbf{G}_k = \begin{bmatrix} \mathbf{G}_k^{vv} & \mathbf{G}_k^{hv} \\ \mathbf{G}_k^{vh} & \mathbf{G}_k^{hh} \end{bmatrix}, \quad (4)$$

in which $\mathbf{G}_k^{mn} \in \mathbb{C}^{r_k \times \frac{N_t}{2}}$ for $m, n \in \{h, v\}$, while the matrix \mathbf{X} is introduced to indicate the power imbalance between both polarizations and it is expressed as

$$\mathbf{X} = \begin{bmatrix} 1 & \sqrt{\chi} \\ \sqrt{\chi} & 1 \end{bmatrix}, \quad (5)$$

¹In fact, our design is equally applicable where the spatial covariance matrices are different in the two polarizations.

² \mathbf{R}_k^p includes the effects of both the antenna correlations at the BS side and at the MU side.

in which $0 \leq \chi \leq 1$ is the inverse of the cross-polarization discrimination (XPD) metric representing the channel's ability to separate vertical and horizontal polarizations [33]. Thus the polarized channel modeling (3) can be rewritten as

$$\begin{aligned} \mathbf{H}_k &= \left(\mathbf{I}_2 \otimes \left(\mathbf{U}_k \mathbf{\Lambda}_k^{\frac{1}{2}} \right) \right) \begin{bmatrix} \mathbf{G}_k^{vv} & \sqrt{\chi} \mathbf{G}_k^{hv} \\ \sqrt{\chi} \mathbf{G}_k^{vh} & \mathbf{G}_k^{hh} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{U}_k \mathbf{\Lambda}_k^{\frac{1}{2}} \mathbf{G}_k^{vv} & \sqrt{\chi} \mathbf{U}_k \mathbf{\Lambda}_k^{\frac{1}{2}} \mathbf{G}_k^{hv} \\ \sqrt{\chi} \mathbf{U}_k \mathbf{\Lambda}_k^{\frac{1}{2}} \mathbf{G}_k^{vh} & \mathbf{U}_k \mathbf{\Lambda}_k^{\frac{1}{2}} \mathbf{G}_k^{hh} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_k^{vv} & \mathbf{H}_k^{hv} \\ \mathbf{H}_k^{vh} & \mathbf{H}_k^{hh} \end{bmatrix}, \end{aligned} \quad (6)$$

and the covariance matrix of \mathbf{H}_k is given by

$$\mathbf{R}_k = \begin{bmatrix} (1 + \chi) \mathbf{R}_k^p & \mathbf{0} \\ \mathbf{0} & (1 + \chi) \mathbf{R}_k^p \end{bmatrix}. \quad (7)$$

Similar to (1), we can formulate the wiretapped signal vector $\mathbf{y}_e \in \mathbb{C}^{N_e}$ at the eavesdropper as

$$\mathbf{y}_e = \mathbf{H}_e^H \mathbf{x} + \mathbf{n}_e, \quad (8)$$

where $\mathbf{n}_e \sim \mathcal{CN}(\mathbf{0}, \sigma_e^2 \mathbf{I}_{N_e})$ is the AWGN vector at eavesdropper and $\mathbf{H}_e \in \mathbb{C}^{N_t \times N_e}$ is the channel matrix from the BS to eavesdropper with the same form as \mathbf{H}_k , namely,

$$\begin{aligned} \mathbf{H}_e &= \left(\mathbf{I}_2 \otimes \left(\mathbf{U}_e \mathbf{\Lambda}_e^{\frac{1}{2}} \right) \right) \begin{bmatrix} \mathbf{G}_e^{vv} & \sqrt{\chi} \mathbf{G}_e^{hv} \\ \sqrt{\chi} \mathbf{G}_e^{vh} & \mathbf{G}_e^{hh} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{U}_e \mathbf{\Lambda}_e^{\frac{1}{2}} \mathbf{G}_e^{vv} & \sqrt{\chi} \mathbf{U}_e \mathbf{\Lambda}_e^{\frac{1}{2}} \mathbf{G}_e^{hv} \\ \sqrt{\chi} \mathbf{U}_e \mathbf{\Lambda}_e^{\frac{1}{2}} \mathbf{G}_e^{vh} & \mathbf{U}_e \mathbf{\Lambda}_e^{\frac{1}{2}} \mathbf{G}_e^{hh} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_e^{vv} & \mathbf{H}_e^{hv} \\ \mathbf{H}_e^{vh} & \mathbf{H}_e^{hh} \end{bmatrix}, \end{aligned} \quad (9)$$

in which $\mathbf{\Lambda}_e \in \mathbb{C}^{r_e \times r_e}$ is the diagonal matrix whose diagonal elements are the r_e non-zero eigenvalues of the eavesdropper's polarized spatial covariance matrix \mathbf{R}_e^p , and the unitary matrix $\mathbf{U}_e \in \mathbb{C}^{\frac{N_e}{2} \times r_e}$ consists of the corresponding eigenvectors of \mathbf{R}_e^p , while the elements of $\mathbf{G}_e^{mn} \in \mathbb{C}^{r_e \times \frac{N_e}{2}}$, $m, n \in \{h, v\}$, follow the circularly symmetric Gaussian distribution with zero mean and unit power, similar to \mathbf{G}_k^{mn} .

B. Optimal linear precoding

Generally, the metric called *achievable secrecy rate* [33] is adopted to measure the security performance of the dual-polarized DL MIMO system, which is expressed as

$$R_{\text{sec}} = \left[\sum_{k=1}^K I_k(s_k, y_k) - I_e(s_1, \dots, s_K, y_e) \right]^+, \quad (10)$$

where $I_k(s_k, y_k)$ and $I_e(s_1, \dots, s_K, y_e)$ denote the mutual information between the BS and k th MU and between the BS and eavesdropper, respectively, which can be expressed as

$$I_k(s_k, y_k) = \log \det \left(\mathbf{I}_{N_r} + \frac{\mathbf{H}_k^H \mathbf{V}_k \mathbf{V}_k^H \mathbf{H}_k}{\sigma_k^2 \mathbf{I}_{N_r} + \sum_{j \neq k} \mathbf{H}_j^H \mathbf{V}_j \mathbf{V}_j^H \mathbf{H}_j} \right), \quad (11)$$

$$I_e(s_1, \dots, s_K, y_e) = \log \det \left(\mathbf{I}_{N_e} + \frac{1}{\sigma_e^2} \sum_{k=1}^K \mathbf{H}_e^H \mathbf{V}_k \mathbf{V}_k^H \mathbf{H}_e \right), \quad (12)$$

assuming Gaussian channels and stochastic encoders.

The optimal linear precoding design should maximize the secrecy rate R_{sec} subject to the system's power constraint. This leads to the following secrecy rate maximization problem

$$\begin{aligned} \max_{\mathbf{V}_k} \quad & \log \left(\frac{\det \left(\mathbf{I}_{N_r} + \frac{\mathbf{H}_k^H \mathbf{V}_k \mathbf{V}_k^H \mathbf{H}_k}{\sigma_k^2 \mathbf{I}_{N_r} + \sum_{j \neq k} \mathbf{H}_j^H \mathbf{V}_j \mathbf{V}_j^H \mathbf{H}_j} \right)}{\det \left(\mathbf{I}_{N_e} + \frac{1}{\sigma_e^2} \sum_{k=1}^K \mathbf{H}_e^H \mathbf{V}_k \mathbf{V}_k^H \mathbf{H}_e \right)} \right), \quad (13) \\ \text{s.t.} \quad & \sum_{k=1}^K \text{Tr}(\mathbf{V}_k \mathbf{V}_k^H) \leq P_{\text{max}}. \end{aligned}$$

Unfortunately, the objective function in (13) consists of multiple generalized Rayleigh quotients. Therefore, this optimization is nonconvex and it is difficult to solve. Although numerical methods, such as Newton method, can be utilized to search for the optimal solutions, the associate computational complexity is huge and the convergence to an optimum cannot be guaranteed. In order to reduce the system feedback cost and the optimization complexity jointly, a specially structured linear precoding scheme is adopted to exploit different availability levels of the eavesdropper's CSI.

C. Dual-structured linear precoding

The dual-structured linear precoding scheme [10] is based on the spatial covariance matrix, i.e., the long-term CSI, and the instantaneous CSI jointly. In particular, the precoding matrix \mathbf{V}_k for the k th MU can be formulated as

$$\mathbf{V}_k = \mathbf{B}_k \mathbf{P}_k, \quad k = 1, 2, \dots, K, \quad (14)$$

where $\mathbf{B}_k \in \mathbb{C}^{N_t \times L}$ is a long-term CSI based preprocessing matrix and $\mathbf{P}_k \in \mathbb{C}^{L \times d_s}$ is the linear precoding matrix based on the instantaneous CSI $\mathbf{H}_k^H \mathbf{B}_k$. The design parameter L determines the transformed channel dimension via the long-term CSI, and generally L satisfies $d_s \leq L \leq 2r_k < N_t$.

The preprocessing matrix \mathbf{B}_k is optimized by only utilizing the long-term CSI, which is generally slowly varying and can be accurately obtained with the low feedback overhead. Therefore, we can firstly design the long-term CSI based preprocessing matrix \mathbf{B}_k , and then apply it to the following design of \mathbf{P}_k . Note that with the block diagonal structure of the channel covariance matrix \mathbf{R}_k , the block diagonalization is also applied to the optimization of \mathbf{B}_k .

Specifically, in order to avoid both the interference to other non-intended MUs and the information leakage to the eavesdropper, the optimal \mathbf{B}_k , $1 \leq k \leq K$, needs to be designed to meet the following condition

$$\mathbf{H}_j^H \mathbf{B}_k \approx \mathbf{0}, \quad \forall j \neq k, \quad \mathbf{H}_e^H \mathbf{B}_k \approx \mathbf{0}. \quad (15)$$

To obtain \mathbf{B}_k that satisfies the above condition, we can utilize the block diagonal structure of \mathbf{R}_k in (7). First, for the k th MU, we define the matrix $\mathbf{U}_{-k} \in \mathbb{C}^{\frac{N_t}{2} \times \left(\sum_{j \neq k} r_j^a + r_e \right)}$ as

$$\mathbf{U}_{-k} = [\tilde{\mathbf{U}}_1 \cdots \tilde{\mathbf{U}}_{k-1} \quad \mathbf{U}_e \quad \tilde{\mathbf{U}}_{k+1} \cdots \tilde{\mathbf{U}}_K], \quad (16)$$

where $\tilde{\mathbf{U}}_k = [\mathbf{U}_k]_{1:r_k^a}$, $r_k^a \leq r_k$, and r_k^a is a design parameter determining the number of dominant eigenvalues of \mathbf{R}_k^p , while the entire \mathbf{U}_e is used in \mathbf{U}_{-k} . By deriving the orthogonal space of \mathbf{U}_{-k} , the complete information leakage from the k th MU to the eavesdropper and the partial interference from the k th

MU to other non-intended MUs can both be suppressed. Thus, the singular value decomposition (SVD) of \mathbf{U}_{-k} is firstly performed to find the orthogonal space of \mathbf{U}_{-k} , that is,

$$\mathbf{U}_{-k} = [\mathbf{Q}_{-k}^0 \ \mathbf{Q}_{-k}^1] \begin{bmatrix} \mathbf{\Lambda}_{-k}^0 & \\ & \mathbf{\Lambda}_{-k}^1 \end{bmatrix} \mathbf{V}_{-k}^H \quad (17)$$

where the diagonal matrices $\mathbf{\Lambda}_{-k}^0$ and $\mathbf{\Lambda}_{-k}^1$ contain the $\sum_{j \neq k} r_j^a + r_e$ dominant singular values (SVs) and the $\frac{N_t}{2} - (\sum_{j \neq k} r_j^a + r_e)$ non-dominant SVs of \mathbf{U}_{-k} , respectively, \mathbf{Q}_{-k}^0 and \mathbf{Q}_{-k}^1 compose of the left singular vectors corresponding to the SVs contained in $\mathbf{\Lambda}_{-k}^0$ and $\mathbf{\Lambda}_{-k}^1$, respectively, while \mathbf{V}_{-k} consists of the right singular vectors of \mathbf{U}_{-k} . Naturally, we can obtain the orthogonal space of \mathbf{U}_{-k} as $\mathbf{Q}_{-k}^1 \in \mathbb{C}^{\frac{N_t}{2} \times (\frac{N_t}{2} - (\sum_{j \neq k} r_j^a + r_e))}$, i.e., $(\mathbf{Q}_{-k}^1)^H \mathbf{U}_{-k} = \mathbf{0}$. Based on this, the optimal structure of \mathbf{B}_k utilizing the block diagonalization is determined as

$$\mathbf{B}_k = \mathbf{I}_2 \otimes \mathbf{B}_k^s, \ \mathbf{B}_k^s = \mathbf{Q}_{-k}^1 \tilde{\mathbf{F}}_k, \quad (18)$$

where $\tilde{\mathbf{F}}_k \in \mathbb{C}^{(\frac{N_t}{2} - (\sum_{j \neq k} r_j^a + r_e)) \times \frac{L}{2}}$ can be derived from the effective channel

$$\tilde{\mathbf{H}}_k = (\mathbf{I}_2 \otimes \mathbf{Q}_{-k}^1)^H \mathbf{H}_k \quad (19)$$

of the k th MU. More specifically, the covariance matrix of $\tilde{\mathbf{H}}_k$ is given by

$$\begin{aligned} \tilde{\mathbf{R}}_k &= (\mathbf{I}_2 \otimes \mathbf{Q}_{-k}^1)^H \mathbf{R}_k (\mathbf{I}_2 \otimes \mathbf{Q}_{-k}^1) \\ &= \begin{bmatrix} (1 + \chi) \tilde{\mathbf{R}}_k^p & \mathbf{0} \\ \mathbf{0} & (1 + \chi) \tilde{\mathbf{R}}_k^p \end{bmatrix}, \end{aligned} \quad (20)$$

where $\tilde{\mathbf{R}}_k^p = (\mathbf{Q}_{-k}^1)^H \mathbf{R}_k^p \mathbf{Q}_{-k}^1$. Then $\tilde{\mathbf{F}}_k$ is determined by the dominant eigenspace of $\tilde{\mathbf{R}}_k^p$. To be specific, the eigenvalue decomposition (EVD) of $\tilde{\mathbf{R}}_k^p$ is given by

$$\tilde{\mathbf{R}}_k^p = \mathbf{F}_k \tilde{\mathbf{\Lambda}}_k \mathbf{F}_k^H, \quad (21)$$

where \mathbf{F}_k consists of the $\min \left\{ \frac{N_t}{2} - (\sum_{j \neq k} r_j^a + r_e), r_k \right\}$ dominant eigenvectors associated with $\tilde{\mathbf{R}}_k^p$. As a result, $\tilde{\mathbf{F}}_k = [\mathbf{F}_k]_{1:\frac{L}{2}}$ is obtained which is then substituted into (18) to obtain the optimal preprocessing matrix \mathbf{B}_k , denoted as \mathbf{B}_k^* .

It is clear that through this preprocessing, we can project the transmit signal for the k th MU onto the L -dimensional dominant eigenspace, which is orthogonal to the partial channel eigenspaces of other non-intended MUs and the entire channel eigenspace of the eavesdropper, that is, $\mathbf{H}_k^H \mathbf{B}_k^* \approx \mathbf{0}$ when r_k^a is approximately equal to r_k and $\mathbf{H}_e^H \mathbf{B}_k^* = \mathbf{0}$, $\forall j, k = 1, \dots, K$ and $j \neq k$. As the rank of $\tilde{\mathbf{R}}_k^p$ is $\text{rank}(\tilde{\mathbf{R}}_k^p) = \min \left\{ \frac{N_t}{2} - (\sum_{j \neq k} r_j^a + r_e), r_k \right\}$, the values of r_k^a and L should be chosen to satisfy the constraint

$$d_s \leq L \leq \min \left\{ N_t - 2 \left(\sum_{j \neq k} r_j^a + r_e \right), 2r_k \right\}, \ \forall k. \quad (22)$$

For brevity, we assume $r_1^a = \dots = r_K^a = r$. The choice of r determines the trade-off between the dimension of the orthogonal space \mathbf{Q}_{-k}^1 and the design freedom of the precoding matrix $\mathbf{P}_k \in \mathbb{C}^{L \times d_s}$. To be specific, if we choose r close to r_k , $\forall k$, the more perfect orthogonal space for the k th MU can

be found, which means that more interference to other non-intended MUs can be suppressed. However, it is clear from (22) that with the increase of r , the range for L decreases. As a result, the design freedom of \mathbf{P}_k is reduced significantly, which may degrade the system performance. Thus, the value of r has important influence on the performance of the dual-polarized DL MIMO system, specifically, the achievable secrecy rate and MSE. In addition, it can be seen that the design of \mathbf{P}_k for $1 \leq k \leq K$ mainly depends on the instantaneous CSIs of the k th MU and the eavesdropper.

In most practical communication scenarios, the eavesdropper is generally hidden from the BS, and its CSI is unavailable. When \mathbf{U}_e is unavailable, we cannot design the precoding matrix \mathbf{B}_k according to the above analysis. Instead, we can only design the precoding matrix, denoted as $\tilde{\mathbf{B}}_k$, based on

$$\tilde{\mathbf{U}}_{-k} = [\tilde{\mathbf{U}}_1 \cdots \tilde{\mathbf{U}}_{k-1} \ \tilde{\mathbf{U}}_{k+1} \cdots \tilde{\mathbf{U}}_K] \in \mathbb{C}^{\frac{N_t}{2} \times \sum_{j \neq k} r_j^a}, \quad (23)$$

in a similar way. Clearly, such a design can only achieve $\mathbf{H}_j^H \tilde{\mathbf{B}}_k \approx \mathbf{0}$, $\forall j \neq k$, and it cannot suppress the information leakage to the eavesdropper.

III. DUAL-STRUCTURED SECURE LINEAR PRECODING DESIGN BASED ON PERFECT CSI

We assume that the perfect CSIs between the BS and all legitimate MUs are available. As for the eavesdropper's CSI, the two cases are discussed. In the first case, the eavesdropper is a legitimate and active but non-intended receiver. In this case, the eavesdropper's CSI can be obtained via a standard training-based channel estimation technique. The precoding matrices \mathbf{B}_k^* , $\forall k$, designed in Section II-C, are capable of suppressing the information leakage to the eavesdropper. In the second case, the eavesdropper's CSI is unavailable to the BS. In this case, the precoding matrices $\tilde{\mathbf{B}}_k$, $\forall k$, cannot block the the information leakage to the eavesdropper, and an artificial noise scheme is applied to resolve the problem. We now detail the linear precoding designs for the both cases.

A. Perfect \mathbf{H}_k and \mathbf{H}_e

Since \mathbf{H}_e is known, the optimal preprocessing matrix \mathbf{B}_k^* can be applied to completely cancel the information leakage to the eavesdropper. Consequently, the received signal (1) at the k th MU can be rewritten as

$$\mathbf{y}_k = \bar{\mathbf{H}}_{kk}^H \mathbf{P}_k \mathbf{s}_k + \sum_{j \neq k} \bar{\mathbf{H}}_{kj}^H \mathbf{P}_j \mathbf{s}_j + \mathbf{n}_k, \quad (24)$$

where $\bar{\mathbf{H}}_{kj} = (\mathbf{B}_j^*)^H \mathbf{H}_k$, $1 \leq k, j \leq K$, while the wiretapped signal of (8) becomes

$$\mathbf{y}_e = \mathbf{n}_e. \quad (25)$$

Therefore, by defining $\bar{\mathbf{B}}_k = (\mathbf{B}_k^*)^H \mathbf{B}_k^* \in \mathbb{C}^{L \times L}$, the secrecy rate maximization problem (13) is transformed into

$$\begin{aligned} \max_{\mathbf{P}_k} \quad & \sum_{k=1}^K \log \det \left(\mathbf{I}_{N_r} + \frac{\bar{\mathbf{H}}_{kk}^H \mathbf{P}_k \mathbf{P}_k^H \bar{\mathbf{H}}_{kk}}{\sigma_k^2 \mathbf{I}_{N_r} + \sum_{j \neq k} \bar{\mathbf{H}}_{kj}^H \mathbf{P}_j \mathbf{P}_j^H \bar{\mathbf{H}}_{kj}} \right), \\ \text{s.t.} \quad & \sum_{k=1}^K \text{Tr}(\mathbf{P}_k^H \bar{\mathbf{B}}_k \mathbf{P}_k) \leq P_{\max}. \end{aligned} \quad (26)$$

To tackle the above nonconvex problem, a WMMSE transformation based BCD algorithm is proposed. More specifically,

through the WMMSE transformation, the problem (26) can be transformed into an equivalent problem by introducing other variables. First, we have the following lemma.

Lemma 1. Define a matrix function $\mathbf{E}(\mathbf{N}, \mathbf{P}) \in \mathbb{C}^{d_s \times d_s}$ for any given $\mathbf{H} \in \mathbb{C}^{L \times N_r}$, $\mathbf{D} \in \mathbb{C}^{N_r \times N_r}$ and $\mathbf{D} \succ \mathbf{0}$ as

$$\mathbf{E}(\mathbf{N}, \mathbf{P}) = (\mathbf{I}_{d_s} - \mathbf{N}^H \mathbf{H}^H \mathbf{P}) (\mathbf{I}_{d_s} - \mathbf{N}^H \mathbf{H}^H \mathbf{P})^H + \mathbf{N}^H \mathbf{D} \mathbf{N}, \quad (27)$$

where $\mathbf{N} \in \mathbb{C}^{N_r \times d_s}$ and $\mathbf{P} \in \mathbb{C}^{L \times d_s}$. Then the following equality holds

$$\begin{aligned} & \log \det(\mathbf{I}_{N_r} + \mathbf{H}^H \mathbf{P} \mathbf{P}^H \mathbf{H} \mathbf{D}^{-1}) \\ &= \max_{\mathbf{W} \succ \mathbf{0}, \mathbf{N}} \log \det(\mathbf{W}) - \text{Tr}(\mathbf{W} \mathbf{E}(\mathbf{N}, \mathbf{P})) + d_s, \end{aligned} \quad (28)$$

where $\mathbf{W} \in \mathbb{C}^{d_s \times d_s}$.

Proof: See Appendix A.

Applying **Lemma 1** with

$$\mathbf{H} = \bar{\mathbf{H}}_{kk}, \quad (29)$$

$$\mathbf{D} = \sigma_k^2 \mathbf{I}_{N_r} + \sum_{j \neq k} \bar{\mathbf{H}}_{kj}^H \mathbf{P}_j \mathbf{P}_j^H \bar{\mathbf{H}}_{kj}, \quad (30)$$

we derive the equivalent optimization problem of (26) as

$$\begin{aligned} & \max_{\mathbf{N}_k, \mathbf{W}_k \succ \mathbf{0}, \mathbf{P}_k} \sum_{k=1}^K \log \det(\mathbf{W}_k) - \text{Tr}(\mathbf{W}_k \mathbf{E}(\mathbf{N}_k, \mathbf{P}_k)) \\ & \quad + d_s, \\ \text{s.t.} \quad & \sum_{k=1}^K \text{Tr}(\mathbf{P}_k^H \bar{\mathbf{B}}_k \mathbf{P}_k) \leq P_{\max}, \end{aligned} \quad (31)$$

where

$$\begin{aligned} \mathbf{E}(\mathbf{N}_k, \mathbf{P}_k) &= (\mathbf{I}_{d_s} - \mathbf{N}_k^H \bar{\mathbf{H}}_{kk}^H \mathbf{P}_k) (\mathbf{I}_{d_s} - \mathbf{N}_k^H \bar{\mathbf{H}}_{kk}^H \mathbf{P}_k)^H \\ & \quad + \mathbf{N}_k^H \left(\sigma_k^2 \mathbf{I}_{N_r} + \sum_{j \neq k} \bar{\mathbf{H}}_{kj}^H \mathbf{P}_j \mathbf{P}_j^H \bar{\mathbf{H}}_{kj} \right) \mathbf{N}_k. \end{aligned} \quad (32)$$

The BCD method can be utilized to iteratively solve the problem (31) by optimizing the objective function of (31) over one set of variables while keeping others invariant at each iteration. Specifically, with the BCD method, the problem (31) can be decomposed into the following three subproblems.

1) \mathbf{W}_k related subproblem: Given the fixed \mathbf{N}_k and \mathbf{P}_k , the optimization problem (31) can be re-expressed as

$$\max_{\mathbf{W}_k \succ \mathbf{0}} \sum_{k=1}^K \log \det(\mathbf{W}_k) - \text{Tr}(\mathbf{W}_k \mathbf{E}(\mathbf{N}_k, \mathbf{P}_k)) + d_s. \quad (33)$$

From the Karush-Kuhn-Tucker (KKT) condition, the optimal \mathbf{W}_k for given \mathbf{N}_k and \mathbf{P}_k is obtained as

$$\mathbf{W}_k^* = (\mathbf{E}(\mathbf{N}_k, \mathbf{P}_k))^{-1}, \forall k. \quad (34)$$

2) \mathbf{N}_k related subproblem: Similarly, the optimal \mathbf{N}_k given \mathbf{P}_k (and \mathbf{W}_k) is obtained by solving the following problem

$$\min_{\mathbf{N}_k} \sum_{k=1}^K \text{Tr}(\mathbf{W}_k \mathbf{E}(\mathbf{N}_k, \mathbf{P}_k)). \quad (35)$$

By substituting $\mathbf{E}(\mathbf{N}_k, \mathbf{P}_k)$ of (32) into the above problem and setting the gradient of the objective function with respect to \mathbf{N}_k to 0, we have the optimal \mathbf{N}_k

$$\mathbf{N}_k^* = \left(\sigma_k^2 \mathbf{I}_{N_r} + \sum_{j=1}^K \bar{\mathbf{H}}_{kj}^H \mathbf{P}_j \mathbf{P}_j^H \bar{\mathbf{H}}_{kj} \right)^{-1} \bar{\mathbf{H}}_{kk}^H \mathbf{P}_k, \quad (36)$$

which turns out to only depend on \mathbf{P}_k .

3) \mathbf{P}_k related subproblem: The \mathbf{P}_k related optimization problem given \mathbf{N}_k and \mathbf{W}_k is formulated by substituting $\mathbf{E}(\mathbf{N}_k, \mathbf{P}_k)$ of (32) into (31), which is

$$\begin{aligned} & \min_{\mathbf{P}_k, \forall k} \sum_{k=1}^K \text{Tr} \left(\mathbf{P}_k^H \sum_{j=1}^K (\bar{\mathbf{H}}_{kj} \mathbf{N}_k \mathbf{W}_k \mathbf{N}_k^H \bar{\mathbf{H}}_{kj}^H) \mathbf{P}_k \right) \\ & \quad - \text{Tr}(\mathbf{W}_k \mathbf{N}_k^H \bar{\mathbf{H}}_{kk}^H \mathbf{P}_k) - \text{Tr}(\mathbf{W}_k \mathbf{P}_k^H \bar{\mathbf{H}}_{kk} \mathbf{N}_k), \\ \text{s.t.} \quad & \sum_{k=1}^K \text{Tr}(\mathbf{P}_k^H \bar{\mathbf{B}}_k \mathbf{P}_k) \leq P_{\max}. \end{aligned} \quad (37)$$

After some manipulations, the objective function of the problem (37) is transformed into a sum of the K \mathbf{P}_k -related functions and the problem (37) can be rewritten as

$$\begin{aligned} & \min_{\mathbf{P}_k, \forall k} \sum_{k=1}^K \text{Tr} \left(\mathbf{P}_k^H \left(\sum_{j=1}^K \bar{\mathbf{H}}_{jk} \mathbf{N}_j \mathbf{W}_j \mathbf{N}_j^H \bar{\mathbf{H}}_{jk}^H \right) \mathbf{P}_k \right) \\ & \quad - \text{Tr}(\mathbf{W}_k \mathbf{N}_k^H \bar{\mathbf{H}}_{kk}^H \mathbf{P}_k) - \text{Tr}(\mathbf{W}_k \mathbf{P}_k^H \bar{\mathbf{H}}_{kk} \mathbf{N}_k), \\ \text{s.t.} \quad & \sum_{k=1}^K \text{Tr}(\mathbf{P}_k^H \bar{\mathbf{B}}_k \mathbf{P}_k) \leq P_{\max}. \end{aligned} \quad (38)$$

The above problem is a standard convex problem, whose Lagrangian is given by

$$\begin{aligned} \mathcal{L}(\mathbf{P}_k, \forall k; \lambda) &= \sum_{k=1}^K \text{Tr} \left(\mathbf{P}_k^H \left(\sum_{j=1}^K \bar{\mathbf{H}}_{jk} \mathbf{N}_j \mathbf{W}_j \mathbf{N}_j^H \bar{\mathbf{H}}_{jk}^H \right) \mathbf{P}_k \right) \\ & \quad - \text{Tr}(\mathbf{W}_k \mathbf{N}_k^H \bar{\mathbf{H}}_{kk}^H \mathbf{P}_k) - \text{Tr}(\mathbf{W}_k \mathbf{P}_k^H \bar{\mathbf{H}}_{kk} \mathbf{N}_k) \\ & \quad + \lambda \left(\sum_{k=1}^K \text{Tr}(\mathbf{P}_k^H \bar{\mathbf{B}}_k \mathbf{P}_k) - P_{\max} \right), \end{aligned} \quad (39)$$

where $\lambda \geq 0$ is the Lagrange multiplier. Then the optimal \mathbf{P}_k^* is readily derived as follows

$$\mathbf{P}_k^* = \left(\sum_{j=1}^K \bar{\mathbf{H}}_{jk} \mathbf{N}_j \mathbf{W}_j \mathbf{N}_j^H \bar{\mathbf{H}}_{jk}^H + \lambda^* \bar{\mathbf{B}}_k \right)^{-1} \bar{\mathbf{H}}_{kk} \mathbf{N}_k \mathbf{W}_k, \quad (40)$$

Furthermore, the optimal λ^* satisfies the following complementary slackness condition

$$\lambda^* \left(\sum_{k=1}^K \text{Tr}((\mathbf{P}_k^*)^H \bar{\mathbf{B}}_k \mathbf{P}_k^*) - P_{\max} \right) = 0. \quad (41)$$

Based on (40) and (41), λ^* is calculated in Proposition 1.

Proposition 1. By decomposing the positive definite matrix $\bar{\mathbf{B}}_k$ into $\bar{\mathbf{B}}_k = \bar{\mathbf{B}}_k^{\frac{1}{2}} (\bar{\mathbf{B}}_k^{\frac{1}{2}})^H$, where $\bar{\mathbf{B}}_k^{\frac{1}{2}} \in \mathbb{C}^{L \times L}$ is full-rank, as well as defining $\mathbf{C}_k = \sum_{j=1}^K \bar{\mathbf{H}}_{jk} \mathbf{N}_j \mathbf{W}_j \mathbf{N}_j^H \bar{\mathbf{H}}_{jk}^H$ and $\mathbf{T}_k = \bar{\mathbf{H}}_{kk} \mathbf{N}_k \mathbf{W}_k$, we have the eigen-decomposition $\bar{\mathbf{B}}_k^{-\frac{1}{2}} \mathbf{C}_k (\bar{\mathbf{B}}_k^{-\frac{1}{2}})^H = \mathbf{L}_k \mathbf{\Xi}_k \mathbf{L}_k^H$, in which $\mathbf{\Xi}_k \in \mathbb{C}^{L \times L}$ is the diagonal matrix whose elements are the eigenvalues of $\bar{\mathbf{B}}_k^{-\frac{1}{2}} \mathbf{C}_k (\bar{\mathbf{B}}_k^{-\frac{1}{2}})^H$ and $\mathbf{L}_k \in \mathbb{C}^{L \times L}$ consists of the corresponding orthogonal eigenvectors. Further defining the matrix $\tilde{\mathbf{T}}_k = \mathbf{L}_k^H \bar{\mathbf{B}}_k^{-\frac{1}{2}} \mathbf{T}_k \mathbf{T}_k^H (\bar{\mathbf{B}}_k^{-\frac{1}{2}})^H \mathbf{L}_k$, then the optimal Lagrangian multiplier λ^* satisfying (41) is given by

$$\begin{cases} \lambda^* = 0, & \text{if } \text{Tr}(\tilde{\mathbf{P}}_k^* \bar{\mathbf{B}}_k (\tilde{\mathbf{P}}_k^*)^H) \leq P_{\max}, \\ \lambda^* > 0, & \text{otherwise,} \end{cases} \quad (42)$$

in which $\tilde{\mathbf{P}}_k^* = \left(\sum_{j=1}^K \bar{\mathbf{H}}_{jk} \mathbf{N}_j \mathbf{W}_j \mathbf{N}_j^H \bar{\mathbf{H}}_{jk}^H \right)^\dagger \bar{\mathbf{H}}_{kk} \mathbf{N}_k \mathbf{W}_k$, and the specific value of $\lambda^* > 0$ is obtained by solving

$$\sum_{k=1}^K \sum_{m=1}^L \frac{\tilde{\mathbf{T}}_k|_{m,m}}{(\mathbf{\Xi}_k|_{m,m} + \lambda^*)^2} = P_{\max}. \quad (43)$$

Proof: See Appendix B. ■

Note that $\tilde{T}_k|_{m,m}/(\Xi_k|_{m,m} + \lambda)^2$ is monotonically decreasing for the nonnegative λ . Therefore, a one-dimensional search, e.g., the bisection method, can be applied to solve (43) effectively. Once the optimal λ^* is determined, we can substitute it into (40) to get the optimal \mathbf{P}_k^* for $1 \leq k \leq K$.

The proposed BCD algorithm for solving the secrecy rate maximization problem (26) is summarized in Algorithm 1, where steps 3-5 correspond to the three optimization subproblems, respectively, and the function

$$R(\mathbf{P}_k | k = 1, \dots, K) = \sum_{k=1}^K \log \det \left(\mathbf{I}_{N_r} + \frac{\tilde{\mathbf{H}}_{kk}^H \mathbf{P}_k \mathbf{P}_k^H \tilde{\mathbf{H}}_{kk}}{\sigma_k^2 \mathbf{I}_{N_r} + \sum_{j \neq k} \tilde{\mathbf{H}}_{kj}^H \mathbf{P}_j \mathbf{P}_j^H \tilde{\mathbf{H}}_{kj}} \right), \quad (44)$$

is the secrecy rate. It has been demonstrated in [34] that \mathbf{P}_k obtained by this BCD algorithm is actually a KKT point of the secrecy rate maximization problem (26).

Algorithm 1 The proposed BCD algorithm

Initialize: Give $\mathbf{P}_k \forall k$ that satisfy $\sum_{k=1}^K \text{Tr}(\mathbf{P}_k^H \tilde{\mathbf{B}}_k \mathbf{P}_k) = P_{\max}$, specify the termination threshold ϵ , and set iteration index $t = 0$

1: **repeat**

2: $t = t + 1, \mathbf{P}_k^t \leftarrow \mathbf{P}_k$

3: $\mathbf{N}_k^t \leftarrow \left(\sigma_k^2 \mathbf{I}_{N_r} + \sum_{j=1}^K \tilde{\mathbf{H}}_{kj}^H \mathbf{P}_j^t (\mathbf{P}_j^t)^H \tilde{\mathbf{H}}_{kj} \right)^{-1} \tilde{\mathbf{H}}_{kk}^H \mathbf{P}_k^t$

4: $\mathbf{W}_k^t \leftarrow (\mathbf{E}(\mathbf{N}_k^t, \mathbf{P}_k^t))^{-1}$

5: With \mathbf{N}_k^t and \mathbf{W}_k^t , find λ^* by solving (43) using Algorithm 2, and update \mathbf{P}_k using (40)

6: **until**

$$|R(\mathbf{P}_k^t | k = 1, \dots, K) - R(\mathbf{P}_k | k = 1, \dots, K)| \leq \epsilon$$

Algorithm 2 The bisection method for finding optimal λ^*

Initialize: Give $\lambda_{lw} < \lambda_{up}$, and specify the stopping threshold ϵ

1: **repeat**

2: $\lambda = \frac{\lambda_{lw} + \lambda_{up}}{2}$

3: Calculate $\sum_{k=1}^K \sum_{m=1}^L \frac{\tilde{T}_k|_{m,m}}{(\Xi_k|_{m,m} + \lambda)^2}$

4: **if** $\left(\sum_{k=1}^K \sum_{m=1}^L \frac{\tilde{T}_k|_{m,m}}{(\Xi_k|_{m,m} + \lambda)^2} - P_{\max} \right) \geq 0$ **then**

5: $\lambda_{lw} = \lambda$

6: **else**

7: $\lambda_{up} = \lambda$

8: **end if**

9: **until** $|\lambda_{up} - \lambda_{lw}| \leq \epsilon$

B. Perfect \mathbf{H}_k and unknown \mathbf{H}_e

Since \mathbf{H}_e is unavailable, we can only design the precoding matrices $\tilde{\mathbf{B}}_k \forall k$ based on (23), which cannot block the information leakage to eavesdropper. In order to make the wiretapped information by eavesdropper as little as possible,

an artificial noise scheme is applied, in which the signal transmitted by the BS is given by

$$\mathbf{x} = \sum_{k=1}^K \tilde{\mathbf{B}}_k \mathbf{P}_k \mathbf{s}_k + \mathbf{n}_a. \quad (45)$$

The artificial noise \mathbf{n}_a is given in the form of

$$\mathbf{n}_a = \mathbf{H}_{\perp} \mathbf{z}, \quad (46)$$

where $\mathbf{z} \in \mathbb{C}^{N_t - KN_r}$ obeys $\mathcal{CN}(\mathbf{0}, \sigma_z^2 \mathbf{I}_{N_t - KN_r})$, and the matrix $\mathbf{H}_{\perp} \in \mathbb{C}^{N_t \times (N_t - KN_r)}$ denotes the orthogonal projection matrix onto the null-space of $\mathbf{H} = [\mathbf{H}_1 \mathbf{H}_2 \dots \mathbf{H}_K]^H$. This design ensures that the additional interference to each MU caused by \mathbf{n}_a can be cancelled completely.

The total transmit power is divided into $P_{\max} = P_{IN} + P_{AN}$, where $P_{IN} = \text{Tr}(\mathbf{P}_k^H \tilde{\mathbf{B}}_k^H \tilde{\mathbf{B}}_k \mathbf{P}_k)$ is the allocated power for information transmission and $P_{AN} = \sigma_z^2 (N_t - KN_r)$ is the allocated power for artificial noise transmission. Due to the unknown eavesdropper's CSI, the power optimization between P_{IN} and P_{AN} to maximize the secrecy rate is impossible. Therefore, we choose the MSE of each MU's received signal as the system quality of service (QoS) metric to allow us to minimize the information transmit power P_{IN} , while simultaneously generating more artificial noise to confuse eavesdropper. Specifically, the received signal of the k th MU under the existence of artificial noise is given by

$$\mathbf{y}_k = \sum_{j=1}^K \mathbf{H}_k^H \tilde{\mathbf{B}}_j \mathbf{P}_j \mathbf{s}_j + \tilde{\mathbf{n}}_k, \quad (47)$$

where $\tilde{\mathbf{n}}_k = \mathbf{H}_k^H \mathbf{n}_a + \mathbf{n}_k = \mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \sigma_k^2 \mathbf{I}_{N_r})$, since $\mathbf{H}_k^H \mathbf{H}_{\perp} = \mathbf{0}$. Let $\mathbf{M}_k \in \mathbb{C}^{N_r \times d_s}$ be the receive filter matrix of the k th MU. Then the recovered signal of the k th MU, $\hat{\mathbf{s}}_k = \mathbf{M}_k^H \mathbf{y}_k$, satisfies

$$\begin{aligned} \mathbf{MSE}_k &= E[(\hat{\mathbf{s}}_k - \mathbf{s}_k)(\hat{\mathbf{s}}_k - \mathbf{s}_k)^H] \\ &= (\mathbf{M}_k^H \tilde{\mathbf{H}}_{kk}^H \mathbf{P}_k - \mathbf{I}_{d_s})(\mathbf{M}_k^H \tilde{\mathbf{H}}_{kk}^H \mathbf{P}_k - \mathbf{I}_{d_s})^H \\ &\quad + \sum_{j=1, j \neq k}^K \mathbf{M}_k^H \tilde{\mathbf{H}}_{kj}^H \mathbf{P}_j \mathbf{P}_j^H \tilde{\mathbf{H}}_{kj} \mathbf{M}_k + \sigma_k^2 \mathbf{M}_k^H \mathbf{M}_k \\ &= \mathbf{M}_k^H \left(\sum_{j=1}^K \tilde{\mathbf{H}}_{kj}^H \mathbf{P}_j \mathbf{P}_j^H \tilde{\mathbf{H}}_{kj} + \sigma_k^2 \mathbf{I}_{N_r} \right) \mathbf{M}_k \\ &\quad - \mathbf{M}_k^H \tilde{\mathbf{H}}_{kk}^H \mathbf{P}_k - \mathbf{P}_k^H \tilde{\mathbf{H}}_{kk} \mathbf{M}_k + \mathbf{I}_{d_s}, \end{aligned} \quad (48)$$

where $\tilde{\mathbf{H}}_{kj} = (\tilde{\mathbf{B}}_j)^H \mathbf{H}_k \in \mathbb{C}^{L \times N_r}$, $1 \leq k, j \leq K$. Clearly, $\text{Tr}(\mathbf{MSE}_k)$ is the MSE of the estimator $\hat{\mathbf{s}}_k$.

Based on (48), the security related optimization problem under the unknown eavesdropper's CSI is formulated as

$$\begin{aligned} \min_{\mathbf{P}_k, \mathbf{M}_k} \quad & \sum_{k=1}^K \text{Tr}(\mathbf{P}_k^H \tilde{\mathbf{B}}_k \mathbf{P}_k), \\ \text{s.t.} \quad & \text{Tr}(\mathbf{MSE}_k) \leq \tau_k, \text{ for } k = 1, 2, \dots, K, \end{aligned} \quad (49)$$

where $\tilde{\mathbf{B}}_k = \tilde{\mathbf{B}}_k^H \tilde{\mathbf{B}}_k$, and τ_k is the required MSE threshold for the k th MU. The above problem is nonconvex due to the coupled optimization variables \mathbf{M}_k and \mathbf{P}_k . Hence, we propose an iterative algorithm to optimize \mathbf{M}_k and \mathbf{P}_k separately in each iteration, while keeping the other variable fixed.

1) *Optimizing \mathbf{M}_k :* From (49), it can be seen that only the constraint contains \mathbf{M}_k . Therefore, given \mathbf{P}_k , we take the first derivative of $\text{Tr}(\mathbf{MSE}_k)$ to obtain the optimal \mathbf{M}_k^*

$$\mathbf{M}_k^* = \left(\sigma_k^2 \mathbf{I}_{N_r} + \sum_{j=1}^K \tilde{\mathbf{H}}_{kj}^H \mathbf{P}_j \mathbf{P}_j^H \tilde{\mathbf{H}}_{kj} \right)^{-1} \tilde{\mathbf{H}}_{kk}^H \mathbf{P}_k. \quad (50)$$

2) *Optimizing P_k* : Given M_k , we first define the following auxiliary variables for the optimization of P_k

$$\mathbf{P} = [\mathbf{P}_1^H \ \mathbf{P}_2^H \ \dots \ \mathbf{P}_K^H] \in \mathbb{C}^{d_s \times KL}, \quad (51)$$

$$\tilde{\mathbf{I}} = [\mathbf{I}_{d_s} \ \mathbf{0}_{d_s \times Kd_s}] \in \mathbb{R}^{d_s \times (1+K)d_s}, \quad (52)$$

$$\tilde{\mathbf{M}}_k = \begin{bmatrix} \mathbf{0}_{(k-1)L \times d_s} \\ \mathbf{H}_{kk}^H \mathbf{M}_k \\ \mathbf{0}_{(K-k)L \times d_s} \end{bmatrix} \in \mathbb{C}^{KL \times d_s}, \quad (53)$$

$$\tilde{\mathbf{M}}_k = \text{Bdiag}\{\tilde{\mathbf{H}}_{k1} \mathbf{M}_k, \dots, \tilde{\mathbf{H}}_{kk-1} \mathbf{M}_k, \mathbf{0}_{L \times d_s}, \tilde{\mathbf{H}}_{kk+1} \mathbf{M}_k, \dots, \tilde{\mathbf{H}}_{kK} \mathbf{M}_k\} \in \mathbb{C}^{KL \times Kd_s}, \quad (54)$$

$$\overline{\mathbf{M}}_k = [\tilde{\mathbf{M}}_k \ \tilde{\mathbf{M}}_k] \in \mathbb{C}^{KL \times (K+1)d_s}, \quad (55)$$

$$\tilde{\mathbf{B}} = \text{Bdiag}\{\tilde{\mathbf{B}}_1^H, \tilde{\mathbf{B}}_2^H, \dots, \tilde{\mathbf{B}}_K^H\} \in \mathbb{C}^{KL \times KN_t}. \quad (56)$$

After some manipulations, the problem (49) for optimizing P_k can be equivalently transformed into

$$\begin{aligned} \min_{\mathbf{P}_k} \quad & \|\mathbf{P} \tilde{\mathbf{B}}\|_F, \\ \text{s.t.} \quad & \|\mathbf{P} \overline{\mathbf{M}}_k - \tilde{\mathbf{I}}\|_F \leq \sqrt{\tau_k - \sigma_k^2 \text{Tr}(\mathbf{M}_k \mathbf{M}_k^H)}, \\ & \text{for } k = 1, \dots, K. \end{aligned} \quad (57)$$

The convex problem (57) can be efficiently solved using an interior point method to obtain the optimal P_k for given M_k .

Algorithm 3 The proposed iterative algorithm for solving (49)

Initialize: Given $P_k \ \forall k$ that satisfy $\sum_{k=1}^K \text{Tr}(\mathbf{P}_k^H \hat{\mathbf{B}}_k \mathbf{P}_k) = P_{\max}$, specify the termination threshold ϵ , and set iteration index $t = 0$

1: **repeat**

2: $t = t + 1, \mathbf{P}_k^t \leftarrow \mathbf{P}_k$

3: $\mathbf{M}_k^t \leftarrow \left(\sigma_k^2 \mathbf{I}_{N_r} + \sum_{j=1}^K \tilde{\mathbf{H}}_{kj}^H \mathbf{P}_j^t (\mathbf{P}_j^t)^H \tilde{\mathbf{H}}_{kj} \right)^{-1} \tilde{\mathbf{H}}_{kk}^H \mathbf{P}_k^t$

4: Update P_k by solving the problem (57) using a convex optimization technique

5: **until**

$$\left| \sum_{k=1}^K \text{Tr}((\mathbf{P}_k^t)^H \hat{\mathbf{B}}_k \mathbf{P}_k^t) - \sum_{k=1}^K \text{Tr}(\mathbf{P}_k^H \hat{\mathbf{B}}_k \mathbf{P}_k) \right| \leq \epsilon$$

The proposed iterative algorithm for solving the optimization (49) is summarized in Algorithm 3. At the t th iteration, \mathbf{M}_k^t is updated using (50) with the given \mathbf{P}_k^{t-1} derived from the previous iteration, and then \mathbf{P}_k^t is optimized by solving (57) with the obtained \mathbf{M}_k^t . Hence, \mathbf{P}_k^t is always feasible to the optimization problem (57) at each iteration, which indicates that with the increasing number of iterations, the information transmit power P_{IN} is monotonically decreasing and converges to the minimum value due to the compactness and continuity of the problem (57).

IV. DUAL-STRUCTURED SECURE LINEAR PRECODING DESIGN BASED ON IMPERFECT GLOBAL CSI

The CSI \mathbf{H}_k is specified by the random channel matrix \mathbf{G}_k and the spatial covariance matrix \mathbf{R}_k^p . Generally, \mathbf{R}_k^p changes much more slowly compared to \mathbf{G}_k . Therefore, it is reasonable

to assume that an accurate estimate of \mathbf{R}_k^p can be obtained through a low feedback overhead, and we can focus on the imperfect \mathbf{G}_k , which is expressed mathematically as

$$\mathbf{G}_k \in \mathcal{G}_k = \{\hat{\mathbf{G}}_k + \mathbf{\Delta}_k : \|\mathbf{\Delta}_k\|_F \leq \sigma_{h_k}\}, \quad (58)$$

where $\hat{\mathbf{G}}_k$ is the nominal channel and $\mathbf{\Delta}_k$ is the norm-bounded channel uncertainty [31], [35]. By substituting (58) into the model (3), we further obtain the uncertainty model of \mathbf{H}_k as

$$\mathbf{H}_k \in \mathcal{H}_k = \{\hat{\mathbf{H}}_k + \mathbf{\Delta}'_k : \|\mathbf{\Delta}'_k\|_F \leq \sigma'_{h_k}\}, \quad (59)$$

where $\sigma'_{h_k} = \sqrt{\frac{(1+\chi)\text{Tr}(\mathbf{R}_k^p)}{2r_k}} \sigma_{h_k}$ for $1 \leq k \leq K$. Similarly, for the eavesdropper's CSI \mathbf{H}_e , we have the uncertainty model

$$\mathbf{G}_e \in \mathcal{G}_e = \{\hat{\mathbf{G}}_e + \mathbf{\Delta}_e : \|\mathbf{\Delta}_e\|_F \leq \sigma_{h_e}\}, \quad (60)$$

or

$$\mathbf{H}_e \in \mathcal{H}_e = \{\hat{\mathbf{H}}_e + \mathbf{\Delta}'_e : \|\mathbf{\Delta}'_e\|_F \leq \sigma'_{h_e}\}, \quad (61)$$

where $\sigma'_{h_e} = \sqrt{\frac{(1+\chi)\text{Tr}(\mathbf{R}_e^p)}{2r_e}} \sigma_{h_e}$, while $\hat{\mathbf{G}}_e$ is the eavesdropper's nominal channel and $\mathbf{\Delta}_e$ is its channel uncertainty.

The preprocessing matrix \mathbf{B}_k of the dual-structured precoding scheme is based on the long-term CSIs, \mathbf{R}_k^p and \mathbf{R}_e^p . Therefore, given the accurate estimates of \mathbf{R}_k^p and \mathbf{R}_e^p , the optimal \mathbf{B}_k^* obtained in Section II can be utilized, which completely cancels the information leakage to eavesdropper, even if the instantaneous CSI of eavesdropper is inexact. We further consider the joint optimization of the linear precoding \mathbf{P}_k and the receive filter \mathbf{M}_k to recover the confidential signals effectively, so as to realize more reliable communications. However, owing to the existence of the short-term CSI estimation errors, the worst-case optimization must be considered, where the MSE of each MU received signal and the transmit power at the BS both play important role. It is worth noting that the most unfavourable channel estimation always results in the largest MSE or transmit power. Therefore, the worst-case optimization can be regarded to be the most robust.

More specifically, we aim to minimize the maximum MSE among all the MUs under the case of imperfect instantaneous CSI information to guarantee the secure communications for the weakest link. Mathematically, this is expressed as

$$\begin{aligned} \min_{\mathbf{P}_k, \mathbf{M}_k} \quad & \max_{\mathbf{H}_k \in \mathcal{H}_k, \forall k} \text{Tr}(\mathbf{MSE}_k), \\ \text{s.t.} \quad & \sum_{k=1}^K \text{Tr}(\mathbf{P}_k^H \bar{\mathbf{B}}_k \mathbf{P}_k) \leq P_{\max}, \end{aligned} \quad (62)$$

with $\bar{\mathbf{B}}_k = (\mathbf{B}_k^*)^H \mathbf{B}_k^*$. The problem (62) can be rewritten as

$$\begin{aligned} \min_{\mathbf{P}_k, \mathbf{M}_k, \mathbf{H}_k \in \mathcal{H}_k, \forall k} \quad & \tau, \\ \text{s.t.} \quad & \sum_{k=1}^K \text{Tr}(\mathbf{P}_k^H \bar{\mathbf{B}}_k \mathbf{P}_k) \leq P_{\max}, \\ & \text{Tr}(\mathbf{MSE}_k) \leq \tau, 1 \leq k \leq K. \end{aligned} \quad (63)$$

Further express the two constraint functions of (63) as

$$\sum_{k=1}^K \text{Tr}(\mathbf{P}_k^H \bar{\mathbf{B}}_k \mathbf{P}_k) = \sum_{k=1}^K \|\mathbf{B}_k^* \mathbf{P}_k\|_F^2, \quad (64)$$

$$\begin{aligned} \text{Tr}(\mathbf{MSE}_k) = & \|\mathbf{M}_k^H \mathbf{H}_k^H \mathbf{B}_k^* \mathbf{P}_k - \mathbf{I}_{d_s}\|_F^2 \\ & + \sum_{j=1, j \neq k}^K \|\mathbf{M}_k^H \mathbf{H}_k^H \mathbf{B}_j^* \mathbf{P}_j\|_F^2 + \sigma_k^2 \|\mathbf{M}_k^H\|_F^2, 1 \leq k \leq K. \end{aligned} \quad (65)$$

Clearly $\text{Tr}(\mathbf{MSE}_k)$ is biconvex in \mathbf{P}_k and \mathbf{M}_k , but it is semi-infinite due to the channel uncertainty Δ'_k . We reformulate each MSE term into the form that is affine in Δ'_k . With the aid of linear matrix inequality (LMI) and the sign-definiteness lemma [31], an equivalent SDP transformation to the problem (63) is conducted. This transformed problem is still biconvex in \mathbf{P}_k and \mathbf{M}_k , and we use an iterative algorithm to alternately perform the convex search for \mathbf{P}_k and \mathbf{M}_k , respectively.

Firstly, based on the identity $\|\mathbf{A}\|_F = \|\text{vec}(\mathbf{A})\|$, where $\text{vec}(\bullet)$ is the column stacking operator, $\text{Tr}(\mathbf{MSE}_k)$ can be rewritten as $\text{Tr}(\mathbf{MSE}_k) = \|\mathbf{m}_k\|^2$ by defining

$$\mathbf{m}_k = \begin{bmatrix} \text{vec}(\mathbf{M}_k^H \mathbf{H}_k^H \mathbf{B}_1^* \mathbf{P}_1) \\ \vdots \\ \text{vec}(\mathbf{M}_k^H \mathbf{H}_k^H \mathbf{B}_K^* \mathbf{P}_K) \\ \sigma_k \text{vec}(\mathbf{M}_k^H) \end{bmatrix} - \mathbf{i}_k \in \mathbb{C}^{(Kd_s + N_r)d_s}, \quad (66)$$

$$\mathbf{i}_k = \begin{bmatrix} \mathbf{0}_{(k-1)d_s^2} \\ \text{vec}(\mathbf{I}_{d_s}) \\ \mathbf{0}_{((K-k)d_s + N_r)d_s} \end{bmatrix}. \quad (67)$$

Combining the channel uncertainty model (59) with the identity $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}(\mathbf{B})$, where \mathbf{A} , \mathbf{B} and \mathbf{C} have appropriate dimensions, the channel uncertainty term Δ'_k can be separated out from \mathbf{m}_k . Thus \mathbf{m}_k can be rewritten as

$$\mathbf{m}_k = \widetilde{\mathbf{m}}_k + \sum_{j=1}^K \mathbf{M}_{kj} \text{vec}(\Delta'_k), \quad (68)$$

where

$$\widetilde{\mathbf{m}}_k = \begin{bmatrix} \text{vec}(\mathbf{M}_k^H \widehat{\mathbf{H}}_k^H \mathbf{B}_1^* \mathbf{P}_1) \\ \vdots \\ \text{vec}(\mathbf{M}_k^H \widehat{\mathbf{H}}_k^H \mathbf{B}_K^* \mathbf{P}_K) \\ \sigma_k \text{vec}(\mathbf{M}_k^H) \end{bmatrix} - \mathbf{i}_k, \quad (69)$$

$$\mathbf{M}_{kj} = \begin{bmatrix} \mathbf{0}_{(j-1)d_s^2 \times N_r N_t} \\ ((\mathbf{B}_j^* \mathbf{P}_j)^T \otimes \mathbf{M}_k^H) \\ \mathbf{0}_{d_s((K-j)d_s + N_r) \times N_r N_t} \end{bmatrix} \in \mathbb{C}^{(Kd_s + N_r)d_s \times N_r N_t}. \quad (70)$$

Owing to the fact that $\|\mathbf{m}_k\|^2 \leq \tau$ is constrained in the optimization problem (63), Schur complementary lemma [36] can be applied to form the equivalent LMI, which is

$$\begin{bmatrix} \tau & \mathbf{m}_k^H \\ \mathbf{m}_k & \mathbf{I}_{n_{\text{LMI}}} \end{bmatrix} \succeq \mathbf{0}, \quad (71)$$

with $n_{\text{LMI}} = (Kd_s + N_r)d_s$. Substituting (68) into (71) yields

$$\begin{bmatrix} \tau & \widetilde{\mathbf{m}}_k^H \\ \widetilde{\mathbf{m}}_k & \mathbf{I}_{n_{\text{LMI}}} \end{bmatrix} \succeq \sum_{j=1}^K \begin{bmatrix} 0 & -(\text{vec}(\Delta'_k))^H \mathbf{M}_{kj}^H \\ -\mathbf{M}_{kj} \text{vec}(\Delta'_k) & \mathbf{0} \end{bmatrix}. \quad (72)$$

Although (72) is semi-infinite due to the channel uncertainty term Δ'_k , the sign-definiteness lemma [31], which is introduced in the following lemma, can be applied to transform it into a finite biconvex form.

Lemma 2. *Given the Hermitian matrix \mathbf{Z} and the matrix set $\{\mathbf{P}_j, \mathbf{Q}_j\}_{j=1}^K$, the semi-infinite LMI with the following form*

$$\mathbf{Z} \succeq \sum_{j=1}^K (\mathbf{P}_j^H \mathbf{X}_j \mathbf{Q}_j + \mathbf{Q}_j^H \mathbf{X}_j^H \mathbf{P}_j), \quad \|\mathbf{X}_j\| \leq \psi_j, \quad \forall j \quad (73)$$

holds if and only if there exist the nonnegative real numbers ξ_1, \dots, ξ_K , such that

$$\begin{bmatrix} \mathbf{Z} - \sum_{i=1}^N \xi_i \mathbf{Q}_i^H \mathbf{Q}_i & -\psi_1 \mathbf{P}_1^H & \cdots & -\psi_K \mathbf{P}_K^H \\ -\psi_1 \mathbf{P}_1 & \xi_1 \mathbf{I} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ -\psi_K \mathbf{P}_K & \mathbf{0} & \cdots & \xi_K \mathbf{I} \end{bmatrix} \succeq \mathbf{0}, \quad (74)$$

where all the matrices involved have appropriate dimensions, and the matrix norm $\|\mathbf{X}_j\|$ in (73) is the spectral norm which becomes the Euclidean norm if \mathbf{X}_j is a vector.

In order to apply Lemma 2 to (72), we introduce

$$\mathbf{Z}_k = \begin{bmatrix} \tau & \widetilde{\mathbf{m}}_k^H \\ \widetilde{\mathbf{m}}_k & \mathbf{I}_{n_{\text{LMI}}} \end{bmatrix} \in \mathbb{C}^{(1+n_{\text{LMI}}) \times (1+n_{\text{LMI}})}, \quad (75)$$

$$\mathbf{P}_{kj} = [\mathbf{0}_{N_r N_t} \quad \mathbf{M}_{kj}^H] \in \mathbb{C}^{N_t N_r \times (1+n_{\text{LMI}})}, \quad (76)$$

$$\mathbf{Q}_{kj} = [-1 \quad \mathbf{0}_{n_{\text{LMI}}}^T], \quad (77)$$

$$\mathbf{X}_j = \text{vec}(\Delta'_j) \in \mathbb{C}^{N_t N_r}. \quad (78)$$

Then (72) can be re-casted as

$$\begin{bmatrix} \mathbf{Z}_k - \sum_{j=1}^K \xi_j \mathbf{Q}_{kj}^H \mathbf{Q}_{kj} & -\sigma'_{h_1} \mathbf{P}_{k1}^H & \cdots & -\sigma'_{h_K} \mathbf{P}_{kK}^H \\ -\sigma'_{h_1} \mathbf{P}_{k1} & \xi_1 \mathbf{I}_{N_t N_r} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ -\sigma'_{h_K} \mathbf{P}_{kK} & \mathbf{0} & \cdots & \xi_K \mathbf{I}_{N_t N_r} \end{bmatrix} \succeq \mathbf{0}, \quad (79)$$

where $\xi_j > 0$ for $1 \leq j \leq K$ can always be found according to Lemma 2 such that the LMI (79) holds for any $1 \leq k \leq K$. Thus the equivalent SDP transformation to the problem (63) is obtained as

$$\begin{aligned} \min_{\mathbf{P}_k, \mathbf{M}_k} \quad & \tau, \\ \text{s.t.} \quad & \sum_{k=1}^K \text{Tr}(\mathbf{P}_k^H \bar{\mathbf{B}}_k \mathbf{P}_k) \leq P_{\max}, \\ & \text{LMIs (79) hold for } 1 \leq k \leq K. \end{aligned} \quad (80)$$

To circumvent the nonconvexity of (80), we resort to an iterative ACO algorithm. Specifically, \mathbf{M}_k is first fixed such that the problem (80) becomes a SDP problem in \mathbf{P}_k , which can be efficiently solved. Then the same optimization is performed on \mathbf{M}_k for the obtained \mathbf{P}_k . The iterative procedure is terminated when a desired accuracy is reached. This ACO algorithm is presented in Algorithm 4. Similar to Algorithm 3,

Algorithm 4 The proposed ACO algorithm for solving (63)

Initialize: Give initial $\tau > 0$ and the initial receive filters \mathbf{M}_k for $1 \leq k \leq K$, specify the stopping threshold ϵ , and set the iteration index $t = 0$

1: **repeat**

2: $t = t + 1, \tau^t \leftarrow \tau, \mathbf{M}_k^t \leftarrow \mathbf{M}_k$

3: Solve the problem (80) to obtain the optimal \mathbf{P}_k^t for given \mathbf{M}_k^t and τ^t

4: Solve the problem (80) to obtain the optimal τ and \mathbf{M}_k for the \mathbf{P}_k^t derived from the previous step.

5: **until** $\tau^t - \tau \leq \epsilon$

this ACO algorithm also converges with a decreasing and bounded objective function in the iterative process.

V. SIMULATION RESULTS

Numerical simulations are conducted to evaluate the proposed dual-structured linear precoding scheme for dual-polarized MIMO. The one-ring model is adopted to model the spatial covariance matrix for each MU, where the m th-row and n th-column element of \mathbf{R}_k^p is given by

$$\mathbf{R}_k^p|_{m,n} = \frac{1}{2\Delta_k^a} \int_{-\Delta_k^a}^{\Delta_k^a} e^{-j\pi\lambda_0^{-1}\boldsymbol{\Omega}^T(\alpha+\theta_k)(\mathbf{r}_m-\mathbf{r}_n)} d\alpha, \quad (81)$$

in which λ_0 is the transmit signal wavelength, and $\Delta_k^a \approx \tan^{-1}(p_k/q_k)$ denotes the angular spread of the departure radio frequency (RF) signal to the k th MU with p_k and q_k indicating the radius of the scattering ring and the distance between the BS and the k th MU, respectively. While $\boldsymbol{\Omega}(\alpha) = [\cos(\alpha) \sin(\alpha)]^T$ is the directional vector of the RF signal with the angle of departure α , θ_k is the azimuth direction of the k th MU, and $\mathbf{r}_m = [x_m \ y_m]^T$ is the location vector of the m -th antenna of the dual-polarized array.

In the simulated dual-polarized MIMO DL, the BS is equipped with $N_t = 100$ dual-polarized antennas (50 pairs of horizontally and vertically polarized antennas) to serve the $K = 6$ MUs, each having $N_r = 4$ dual-polarized antennas (2 pairs of horizontally and vertically polarized antennas). The system is wiretapped by an eavesdropper equipped with $N_e = 4$ dual-polarized antennas (2 pairs of horizontally and vertically polarized antennas). The antenna spacing of the dual-polarized array is half of λ_0 and the XPD parameter is defined as $\chi = 0.1$. The angular spreads are $\Delta_1^a = \dots = \Delta_K^a = \frac{4\pi}{180}$, and the azimuth angle of the k th MU is $\theta_k = -\frac{\pi}{6} + \frac{\pi}{3}(k-1)$ for $1 \leq k \leq K$. For the preprocessing matrix $\mathbf{B}_k \in \mathbb{C}^{N_t \times L}$, the value of L is determined according to $L = \min\{N_t - 2((K-1)r + r_e), 2r_k\}$ for $1 \leq k \leq K$, which depends on the chosen value of r . The elements of the instantaneous channels are assumed to follow the distribution $\mathcal{CN}(0, 1)$, and the power of all channel AWGNs is $\sigma_k^2 = \sigma_e^2 = \sigma^2 = 1$. The toolbox CVX [36] is utilized to solve the standard convex optimization problems, e.g., the QCQP of (57) and the SDP of (80).

We also apply the dual-structured linear precoding to the single-polarized MIMO system, i.e., the conventional correlated MIMO. For the sake of fairness, we consider the same numbers of antenna elements in the single-polarized arrays for the BS, MUs and eavesdropper, as in the case of dual-polarized MIMO. The antenna spacing of a single-polarized array is also half of the signal wavelength. Since the size of a single-polarized array is twice of the dual-polarized array with the same number of antenna elements, a single-polarized antenna array may have implementation difficulty, particularly, for a large-scale array. Moreover, the design of the dual-structured linear precoding for single-polarized MIMO imposes much higher complexity than that of the dual-structured linear precoding for dual-polarized MIMO.

As pointed out previously, the existing linear precoding algorithms originally designed for correlated MIMO can also

be applied to dual-polarized MIMO. In the simulation study, we also apply these standard algorithms, specifically, the algorithms of [26], [28], [31], to the dual-polarized MIMO system and use their achievable security performance as the benchmarks. These standard designs are based on the full instantaneous CSI and, therefore, they outperform the dual-structured precoding. However, the full instantaneous CSI is difficult to acquire in practice, and these existing scheme impose huge channel feedback overhead. Moreover, they also impose huge computational burden in designing precoding.

Numerical experiments are performed for the cases of perfect global CSI, completely unknown eavesdropper's CSI and imperfect global CSI, respectively. All the results are averaged over 500 Monte Carlo simulations.

A. The perfect global CSI

In this case, the secrecy rate can be maximized using the proposed BCD algorithm. Given $d_s = 2$ and with $P_{\max} = 16$ dB, Fig. 2 depicts the convergence performance of the BCD algorithm for different initializations and values of L . From Fig. 2, it can be seen that the proposed BCD algorithm guarantees to converge to the optimal secrecy rate. As expected, with the increase of L , the achievable secrecy rate also increases, due to the fact that the expansion of the dominant eigenspace of the effective channel \mathbf{H}_k offers more design freedom for the linear precoding \mathbf{P}_k .

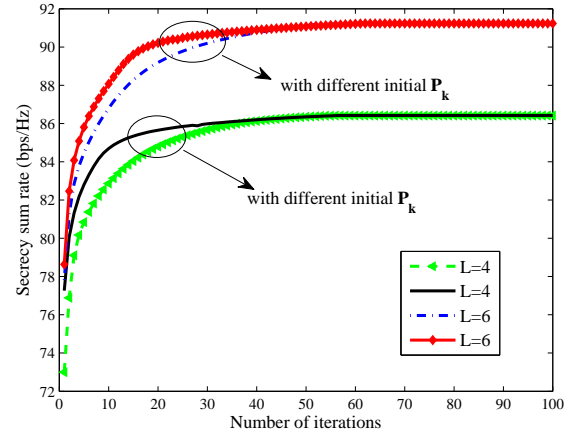


Fig. 2. Convergence performance of the BCD algorithm for different initializations and different values of L , given $d_s = 2$ and $P_{\max} = 16$ dB.

Next, we investigate the influence of the maximum transmit power P_{\max} to the achievable secrecy rates of the three schemes, namely, the dual-structured linear precoding for dual-polarized MIMO, the dual-structured linear precoding for single-polarized MIMO and the null-space design of [26] applied to dual-polarized MIMO. The results of Fig. 3 are for the senario of $L = 4$ and $d_s = 4$, while Fig. 4 depicts the results for the senario of $L = 6$ and $d_s = 4$. As expected, the null-space design [26] outperforms the dual-structured linear precoding design when both applied to the dual-polarized array based MIMO system, while the dual-structured linear precoding for the dual-polarized MIMO DL attains a higher secrecy rate than the dual-structured linear precoding for the single-polarized MIMO DL. Not surprisingly, Figs. 3 and 4

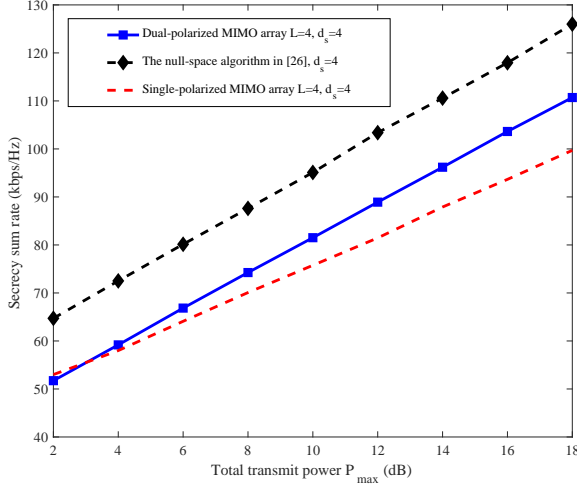


Fig. 3. The achievable secrecy rates as the functions of total transmit power P_{\max} for three different designs, given $L = 4$ and $d_s = 4$.

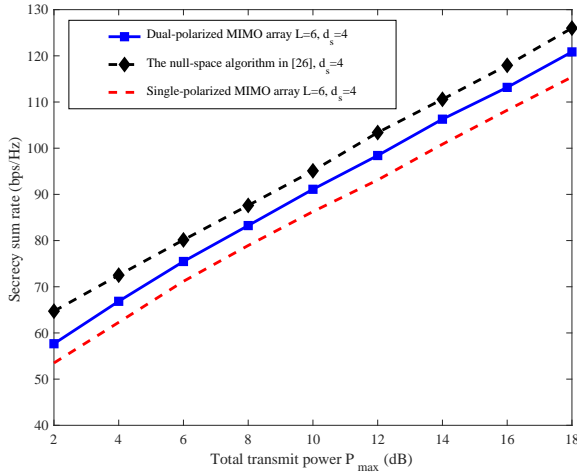


Fig. 4. The achievable secrecy rates as the functions of total transmit power P_{\max} for three different designs, given $L = 6$ and $d_s = 4$.

also indicate that the achievable secrecy sum rates increase with the total transmit power P_{\max} for all the three systems.

The influence of the eavesdropper's wiretap channel \mathbf{H}_e on the secrecy rate performance of both the dual-structured linear

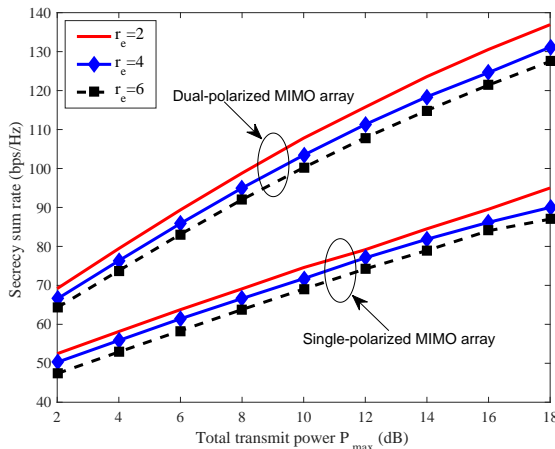


Fig. 5. The achievable secrecy rates as the functions of total transmit power P_{\max} for two different designs under different eavesdropper's channel spatial correlations r_e , given $L = 6$ and $d_s = 2$.

precoding designs for the dual-polarized and single-polarized MIMO DL systems is next investigated. Specifically, in Fig. 5, the achievable secrecy rates are shown as the functions of the total transmit power P_{\max} for the both systems under three different values of r_e . It can be observed from Fig. 5 that for the both designs, the achievable secrecy rates decrease with the increase of r_e . This is because increasing r_e expands the dominant eigenspace of the wiretap channel \mathbf{H}_e , which means that more information leakage to eavesdropper occurs. In addition, the dual-polarized MIMO design always attains a high secrecy rate than the single-polarized MIMO design, as can be evidently seen from Fig. 5.

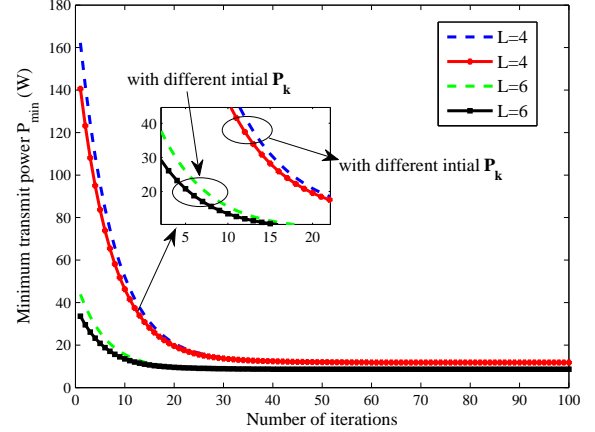


Fig. 6. Convergence performance of Algorithm 3 for different initializations and different values of L , given $d_s = 2$, $P_{\max} = 16$ dB and $\tau_{th} = 0.1$.

B. The completely unavailable eavesdropper's CSI

In this case, an artificial noise scheme is applied to improve the system security performance, and the resulting optimization aims to minimize the information transmit power subject to the prescribed MSE for the recovered confidential signals, which is solved using the iterative algorithm given in Algorithm 3. Given the MSE thresholds $\tau_1 = \dots = \tau_K = \tau_{th} = 0.1$, $P_{\max} = 16$ dB and $d_s = 2$, the convergence performance of this iterative algorithm is depicted in Fig. 6, under different initializations and values of L . It is clear that the proposed algorithm guarantees to converge to the minimum transmit power solution. Moreover, for a given MSE threshold τ_{th} , the minimum total power P_{\min} is reduced with the increase of L owing to the improvement of design freedom for \mathbf{P}_k .

The influence of the MSE threshold τ_{th} on the achievable minimum transmit power P_{\min} for the three schemes, namely, the dual-structured linear precoding for dual-polarized MIMO, the dual-structured linear precoding for single-polarized MIMO and the artificial noise scheme of [28] applied to dual-polarized MIMO, are portrayed in Fig. 7 and Fig. 8, respectively, under two sets of L and d_s values. Clearly, for all the three schemes, the required transmit power P_{\min} is reduced when the MSE requirement is relaxed, i.e., when τ_{th} is increased. The results of Figs. 7 and 8 show that the artificial noise scheme [28] outperforms the dual-structured linear precoding scheme when both are applied to the dual-polarized MIMO, while the dual-structured linear precoding for the dual-polarized MIMO DL achieves a better performance than

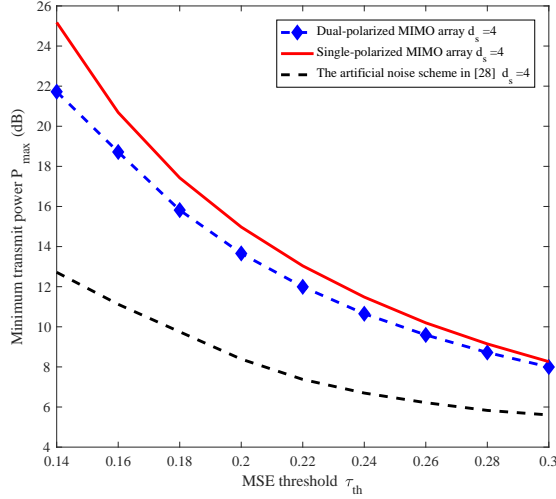


Fig. 7. The achievable minimum transmit powers as the functions of the MSE threshold for three different designs, given $d_s = 4$ and $L = 4$.

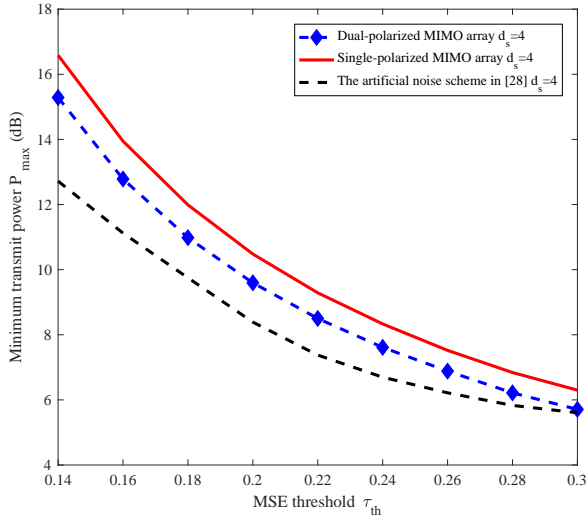


Fig. 8. The achievable minimum transmit powers as the functions of the MSE threshold for three different designs, given $d_s = 4$ and $L = 6$.

the dual-structured linear precoding for the single-polarized MIMO DL, in terms of minimum transmit power. Compared Fig. 8 to Fig. 7, it can be seen that increasing L improves the performance for all the three systems. In particular, we observe that when L is increased from 4 to 6, the minimum transmit power gap between the dual-structured linear precoding and the artificial noise scheme is narrowed significantly.

C. The imperfect global CSI

In this case, we change the number of antennas at the BS to $N_t = 50$. The robust precoding optimization is formulated for achieving secure communications, which is solved using the iterative ACO algorithm presented in Algorithm 4. In the simulation, the channel uncertainty is defined by $\sigma'_{h_1} = \dots = \sigma'_{h_K} = \sigma_h$. Fig. 9 depicts the convergence performance of the ACO algorithm for different initializations and values of L , given $d_s = 2$, $P_{\max} = 16$ dB and $\sigma_h = 0.05$. Similar to Algorithms 1 and 3, the convergence of the iterative ACO algorithm is evident in Fig. 9. Furthermore, the system-wide

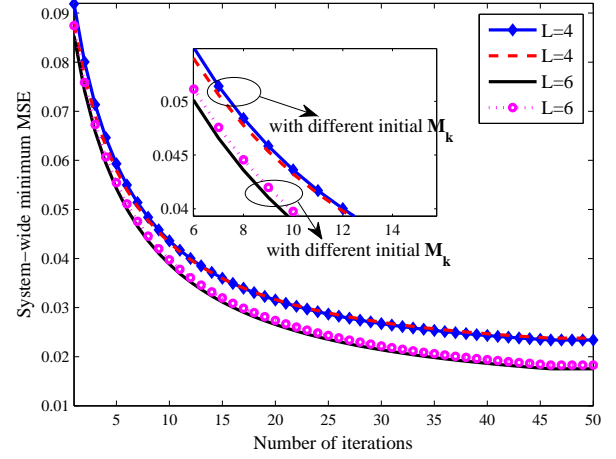


Fig. 9. The convergence performance of Algorithm 4 for different initializations and different values of L , given $d_s = 2$, $P_{\max} = 16$ dB and $\sigma_h = 0.05$.

minimum MSE τ decreases with the increase of L .

Fig. 10 shows the achievable system-wide minimum MSEs as the functions of the total transmit power P_{\max} for the three schemes, namely, the dual-structured linear precoding for dual-polarized MIMO, the dual-structured linear precoding for single-polarized MIMO and the robust optimization algorithm of [31] applied to dual-polarized MIMO, under the condition of $L = 4$, $d_s = 4$ and $\sigma_h = 0.1$. As expected, the achievable system-wide minimum MSE decreases with the increase of the total transmit power in all the three schemes. Not surprisingly, the robust optimization algorithm [31] based on the full instantaneous CSI achieves a lower system-wide minimum MSE than the dual-structured linear precoding, when both are applied to the dual-polarized MIMO DL. It is also clear from Fig. 10 that the dual-polarized MIMO design outperforms the single-polarized MIMO design considerably.

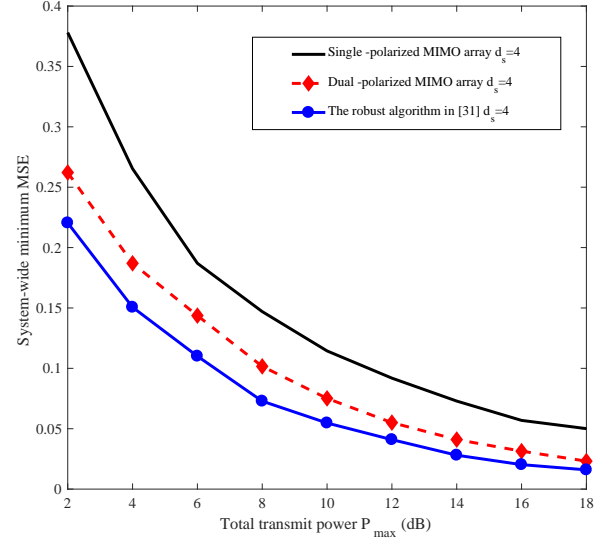


Fig. 10. The system-wide minimum MSEs as the functions of total transmit power for three different designs, given $L = 4$, $d_s = 4$ and $\sigma_h = 0.1$.

Fig. 11 investigates the influence of the channel error σ_h on the achievable system-wide minimum MSE of the dual-structured linear precoding scheme for dual-polarized MIMO DL. It can be observed from Fig. 11 that increasing the

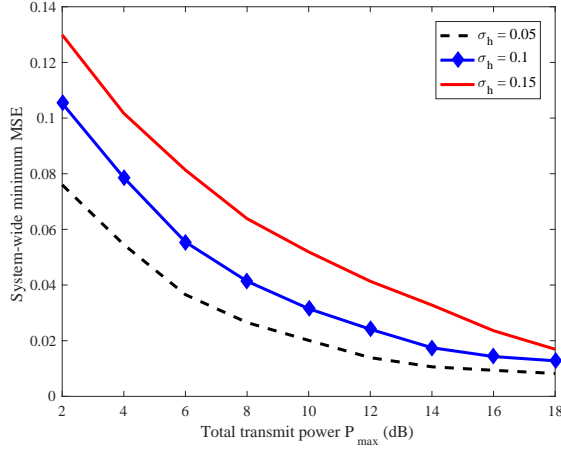


Fig. 11. The system-wide minimum MSE as the function of total transmit power achieved by the dual-structured linear precoding scheme for dual-polarized MIMO under different channel errors σ_h , given $L = 6$ and $d_s = 2$.

channel uncertainty σ_h leads to the increase in the achievable system-wide minimum MSE, simply because the worse CSI information is used in the security performance optimization of the dual-polarized MIMO system.

VI. CONCLUSIONS

We have developed a dual-structured multi-user linear precoding scheme to achieve secure communications for the dual-polarized MIMO system, in which the preprocessing is performed based on the long-term CSI, while the linear precoding is designed based on the instantaneous CSI and the availability of eavesdropper's CSI. Under the ideal scenario of perfect global CSI, the maximum secrecy rate has been attained using the proposed BCD algorithm. For the case of completely unknown eavesdropper's CSI, an artificial noise scheme has been applied to reduce the information leakage to eavesdropper as much as possible, while maintaining the prescribed MSE threshold for the recovered confidential signals. This leads to a security optimization problem that can be efficiently solved by the proposed iterative algorithm. Moreover, considering the practical case of imperfect global CSI, the robust worst-case optimization for achieving secure communications has been formulated, which can be solved using the proposed iterative ACO algorithm. Numerical experiments have been conducted to demonstrate the superior performance of our proposed dual-structured linear precoding designs for dual-polarized MIMO over the dual-structured linear precoding designs for conventional single-polarized MIMO, in terms of the achievable secrecy rate, the minimum transmit power and the MSE performance of recovered confidential signals.

APPENDIX

A. Proof of Lemma 1

Proof: Clearly, the matrix function $\mathbf{E}(\mathbf{N}, \mathbf{P}) \in \mathbb{C}^{d_s \times d_s}$ defined in (27) is positive definite, i.e., $\mathbf{E}(\mathbf{N}, \mathbf{P}) \succ \mathbf{0}$. For any $\mathbf{W} \in \mathbb{C}^{d_s \times d_s}$ and $\mathbf{W} \succ \mathbf{0}$, let us formulate the following joint convex optimization problem in terms of \mathbf{W} and \mathbf{N} as

$$\begin{aligned} \max_{\mathbf{W}, \mathbf{N}} \quad & \log \det(\mathbf{W}) - \text{Tr}(\mathbf{W}\mathbf{E}(\mathbf{N}, \mathbf{P})) + d_s, \\ \text{s.t.} \quad & \mathbf{W} \succ \mathbf{0}. \end{aligned} \quad (82)$$

Applying the KKT condition to the above optimization problem, we have

$$(\mathbf{W}^*)^{-1} - \mathbf{E}(\mathbf{N}^*, \mathbf{P}) - \mathbf{Y}^* = \mathbf{0}, \quad (83)$$

$$\mathbf{Y}^* \mathbf{W}^* = \mathbf{0}, \quad (84)$$

$$\mathbf{H}^H \mathbf{P} \mathbf{P}^H \mathbf{H} \mathbf{N}^* + \mathbf{D} \mathbf{N}^* = \mathbf{H}^H \mathbf{P}, \quad (85)$$

where \mathbf{W}^* and \mathbf{N}^* are the optimal primal variables, while $\mathbf{Y}^* \succeq \mathbf{0}$ is the optimal Lagrangian dual variable corresponding to the constraint of the optimization problem (82). Furthermore, (83) and (85) are the first-order derivatives of the objective function with respect to \mathbf{W} and \mathbf{N} , respectively, while (84) denotes the complementary slackness condition for any positive definite matrix $\mathbf{W} \succ \mathbf{0}$. Considering that $\mathbf{W} \succ \mathbf{0}$ is required, we have the optimal $\mathbf{Y}^* = \mathbf{0}$ and

$$\begin{aligned} \mathbf{W}^* &= (\mathbf{E}(\mathbf{N}^*, \mathbf{P}))^{-1} \\ &= (\mathbf{E}((\mathbf{H}^H \mathbf{P} \mathbf{P}^H \mathbf{H} + \mathbf{D})^{-1} \mathbf{H}^H \mathbf{P}, \mathbf{P}))^{-1} \\ &= (\mathbf{I}_{d_s} - \mathbf{P}^H \mathbf{H} (\mathbf{H}^H \mathbf{P} \mathbf{P}^H \mathbf{H} + \mathbf{D})^{-1} \mathbf{H}^H \mathbf{P})^{-1}. \end{aligned} \quad (86)$$

Substituting (86) into the problem (82), we obtain the optimal objective function value as

$$\begin{aligned} \max_{\mathbf{W} \succ \mathbf{0}, \mathbf{N}} \quad & \log \det(\mathbf{W}) - \text{Tr}(\mathbf{W}\mathbf{E}(\mathbf{N}, \mathbf{P})) + d_s \\ &= \log \det((\mathbf{E}(\mathbf{N}^*, \mathbf{P}))^{-1}) \\ &= \log \det(\mathbf{I}_{d_s} - \mathbf{P}^H \mathbf{H} (\mathbf{H}^H \mathbf{P} \mathbf{P}^H \mathbf{H} + \mathbf{D})^{-1} \mathbf{H}^H \mathbf{P})^{-1} \\ &= \log \det(\mathbf{I}_{d_s} + \mathbf{P}^H \mathbf{H} \mathbf{D}^{-1} \mathbf{H}^H \mathbf{P}), \end{aligned} \quad (87)$$

where the last equality holds due to the identity $(\mathbf{A} + \mathbf{C} \mathbf{B} \mathbf{C}^H)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{C} (\mathbf{B}^{-1} + \mathbf{C}^H \mathbf{A}^{-1} \mathbf{C})^{-1} \mathbf{C}^H \mathbf{A}^{-1}$, with \mathbf{A} , \mathbf{B} and \mathbf{C} having appropriate dimensions. From (87), we arrive at

$$\begin{aligned} & \log \det(\mathbf{I}_{N_r} + \mathbf{H}^H \mathbf{P} \mathbf{P}^H \mathbf{H} \mathbf{D}^{-1}) \\ &= \log \det(\mathbf{I}_{d_s} + \mathbf{P}^H \mathbf{H} \mathbf{D}^{-1} \mathbf{H}^H \mathbf{P}) \\ &= \max_{\mathbf{W} \succ \mathbf{0}, \mathbf{N}} \log \det(\mathbf{W}) - \text{Tr}(\mathbf{W}\mathbf{E}(\mathbf{N}, \mathbf{P})) + d_s, \end{aligned} \quad (88)$$

where the first equality holds based on the identity $\det(\mathbf{I} + \mathbf{A} \mathbf{B}) = \det(\mathbf{I} + \mathbf{B} \mathbf{A})$. ■

B. Proof of Proposition 1

Proof: Obviously, when the power constraint is inactive, i.e., if $\text{Tr}(\tilde{\mathbf{P}}_k^* \tilde{\mathbf{B}}_k (\tilde{\mathbf{P}}_k^*)^H) \leq P_{\max}$, we have $\lambda^* = 0$. Otherwise, we have $\lambda^* > 0$. By the definitions of \mathbf{C}_k and \mathbf{T}_k , the optimal power allocation (40) can be expressed as

$$\mathbf{P}_k^* = (\mathbf{C}_k + \lambda^* \tilde{\mathbf{B}}_k)^{-1} \mathbf{T}_k. \quad (89)$$

Substituting (89) into (41), we have

$$\begin{aligned}
& \text{Tr} \left(\mathbf{T}_k^H (\mathbf{C}_k + \lambda^* \bar{\mathbf{B}}_k)^{-1} \bar{\mathbf{B}}_k (\mathbf{C}_k + \lambda^* \bar{\mathbf{B}}_k)^{-1} \mathbf{T}_k \right) \\
&= \text{Tr} \left(\mathbf{T}_k^H (\bar{\mathbf{B}}_k^{-\frac{1}{2}} \mathbf{C}_k + \lambda^* (\bar{\mathbf{B}}_k^{\frac{1}{2}})^H)^{-1} \right. \\
&\quad \times \left. (\mathbf{C}_k (\bar{\mathbf{B}}_k^{-\frac{1}{2}})^H + \lambda^* \bar{\mathbf{B}}_k^{\frac{1}{2}})^{-1} \mathbf{T}_k \right) \\
&= \text{Tr} \left(\mathbf{T}_k^H (\bar{\mathbf{B}}_k^{-\frac{1}{2}})^H (\bar{\mathbf{B}}_k^{-\frac{1}{2}} \mathbf{C}_k (\bar{\mathbf{B}}_k^{-\frac{1}{2}})^H + \lambda^* \mathbf{I}_L)^{-2} \bar{\mathbf{B}}_k^{-\frac{1}{2}} \mathbf{T}_k \right) \\
&= \text{Tr} \left((\mathbf{L}_k \Xi_k \mathbf{I}_k^H + \lambda^* \mathbf{I}_L)^{-2} \bar{\mathbf{B}}_k^{-\frac{1}{2}} \mathbf{T}_k \mathbf{T}_k^H (\bar{\mathbf{B}}_k^{-\frac{1}{2}})^H \right) \\
&= \text{Tr} \left((\Xi_k + \lambda^* \mathbf{I}_L)^{-2} \tilde{\mathbf{T}}_k \right) = P_{\max}. \tag{90}
\end{aligned}$$

With some mathematical simplifications, we can readily derive (43) from (90). This completes the proof. ■

REFERENCES

- [1] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Euro. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Dec. 1999.
- [2] C. Oestges, V. Erceg, and A. J. Paulraj, "Propagation modeling of MIMO multipolarized fixed wireless channels," *IEEE Trans. Veh. Technol.*, vol. 53, no. 3, pp. 644–654, May 2004.
- [3] H. Weingarten, Y. Steinberg, and S. S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.
- [4] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [5] W. A. Th. Kottterman, G. Sommerkorn, and R. S. Thoma, "Cross-correlation values for dual-polarized indoor MIMO links and realistic antenna elements," in *Proc. 3rd Int. Symp. Wireless Communication Systems* (Valencia, Spain), Sep. 6–8, 2006, pp. 505–509.
- [6] T. Kim, B. Clerckx, D. J. Love, and S. J. Kim, "Limited feedback beamforming systems for dual-polarized MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3425–3439, Nov. 2010.
- [7] L. Dong, H. Choo, R. W. Heath, and H. Ling, "Simulation of MIMO channel capacity with antenna polarization diversity," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1869–1873, Jul. 2005.
- [8] C. Oestges, B. Clerckx, M. Guillaud, and M. Debbah, "Dual-polarized wireless communications: from propagation models to system performance evaluation," *IEEE Trans. Wireless Commun.*, vol. 7, no. 10, pp. 4019–4031, Oct. 2008.
- [9] L. Dai, Z. Gao, and Z. Wang, "Joint channel estimation and feedback with low overhead for FDD massive MIMO systems," in *Proc. IEEE ICC 2015* (Shenzhen, China), Nov. 2–4, 2015, pp. 1–6.
- [10] A. Adhikary, J. Nam, J. Ahn, and G. Caire, "Joint spatial division and multiplexing the large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.
- [11] C. Lim, T. Yoo, B. Clerckx, B. Lee, and B. Shim, "Recent trend of multiuser MIMO in LTE advanced," *IEEE Commun. Mag.*, vol. 51, no. 3, pp. 127–135, Mar. 2013.
- [12] P. Cheng, M. Tao, and W. Zhang, "A new SLNR-based linear precoding for downlink multi-user multi-stream MIMO systems," *IEEE Commun. Lett.*, vol. 14, no. 11, pp. 1008–1010, Nov. 2010.
- [13] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2198–2206, Dec. 2001.
- [14] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Massive MIMO performance evaluation based on measured propagation data," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3899–3911, Jul. 2015.
- [15] J. Park and B. Clerckx, "Multi-user linear precoding for multi-polarized massive MIMO system under imperfect CSIT," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2532–2547, May 2015.
- [16] J. Song, S. G. Larew, D. J. Love, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for multiuser dual-polarized MIMO systems," in *Proc. IEEE GlobalSIP 2013* (Austin, TX), Dec. 3–5, 2013, pp. 719–722.
- [17] M. Bloch, J. Barros, M. R. D. Rodrigues, and S. W. McLaughlin, "Wireless information-theoretic security," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2515–2534, Jun. 2008.
- [18] I. Csiszár and J. Kerner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 339–348, May 1978.
- [19] C. E. Shannon, "Communication theory of secrecy systems," *Bell Syst. Tech. J.*, vol. 28, no. 4, pp. 656–715, Oct. 1949.
- [20] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, Oct. 1975.
- [21] S. K. Leung-Yan-Cheong and M. Hellman, "The Gaussian wire-tap channel," *IEEE Trans. Inf. Theory*, vol. 24, no. 4, pp. 451–456, Jul. 1978.
- [22] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas I: the MISOME wiretap channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3088–3104, Jul. 2010.
- [23] S. Shafiee and S. Ulukus, "Achievable rates in Gaussian MISO channels with secrecy constraints," in *Proc. ISIT 2007* (Nice, France), Jun. 24–29, 2007, pp. 2466–2470.
- [24] G. Pan, C. Tang, T. Li, and Y. Chen, "Secrecy performance analysis for SISO simultaneous wireless information and power transfer systems," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3423–3433, Sep. 2015.
- [25] Y. Wang, F. R. Yu, H. Tang, and M. Huang, "A mean field game theoretic approach for security enhancements in mobile ad hoc networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1616–1627, Mar. 2014.
- [26] H. M. Wang, Q. Yin, and X. G. Xia, "Improving the physical-layer security of wireless two-way relaying via analog network coding," in *Proc. GLOBECOM 2011* (Houston, TX), Dec. 5–9, 2011, pp. 1–6.
- [27] J. Zhu, R. Schober, and V. K. Bhargava, "Secure transmission in multicell massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 4766–4781, Sep. 2014.
- [28] J. Zhu, R. Schober, and V. K. Bhargava, "Linear precoding of data and artificial noise in secure massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 2245–2261, Mar. 2016.
- [29] Y. Ding and V. Fusco, "Polarization distortion as a means for securing wireless communication," in *Proc. EuCAP 2014* (Hague, Netherlands), Apr. 6–11, 2014, pp. 1203–1207.
- [30] B. Clerckx, G. Kim, and S. Kim, "Correlated fading in broadcast MIMO channels: curse or blessing ?" in *Proc. GLOBECOM 2008* (New Orleans, LO), Nov. 30–Dec. 4, 2008, pp. 1–5.
- [31] E. A. Gharavol and E. G. Larsson, "The sign-definiteness lemma and its applications to robust transceiver optimization for multiuser MIMO systems," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 238–252, Jan. 2013.
- [32] J. Park and B. Clerckx, "Multi-user linear precoding for multi-polarized massive MIMO system under imperfect CSIT," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2532–2547, May 2015.
- [33] E. Tekin and A. Yener, "The general Gaussian multiple access and two-way wire-tap channels: achievable rates and cooperative jamming," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2735–2751, Jun. 2008.
- [34] Q. Shi, W. Xu, J. Wu, E. Song, and Y. Wang, "Secure beamforming for MIMO broadcasting with wireless information and power transfer," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2841–2853, May 2015.
- [35] G. Zheng, K.-K. Wong, and B. Ottersten, "Robust cognitive beamforming with bounded channel uncertainties," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4871–4881, Dec. 2009.
- [36] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press: New York, 2004.



Shiqi Gong (S'14) received the B.S. degree in Electronic Engineering in 2014 from Beijing Institute of Technology, Beijing, China. Currently, she is pursuing the Ph.D. degree with the School of Electronic and Information, Beijing Institute of Technology. Her research interests are in the area of signal processing, physical-layer security, resource allocation, and convex optimization.



Chengwen Xing (S'08-M'10) received the B.Eng. degree from Xidian University, Xian, China, in 2005, and the Ph.D. degree from the University of Hong Kong, Hong Kong, China, in 2010.

Since September 2010, he has been with the School of Information and Electronics, Beijing Institute of Technology, Beijing, China, where he is currently an Associate Professor. From September 2012 to December 2012, he was a visiting scholar at the University of Macau. His current research interests include statistical signal processing, convex optimization, multivariate statistics, combinatorial optimization, massive MIMO systems, and high frequency band communication systems. Dr. Xing is an Associate Editor for the IEEE Transactions On Vehicular Technology, KSII Transactions on Internet and Information Systems, Transactions on Emerging Telecommunications Technologies, and China Communications.



Sheng Chen (M'90-SM'97-F'08) received his BEng degree from the East China Petroleum Institute, Dongying, China, in 1982, and his PhD degree from the City University, London, in 1986, both in control engineering. In 2005, he was awarded the higher doctoral degree, Doctor of Sciences (DSc), from the University of Southampton, Southampton, UK.

From 1986 to 1999, He held research and academic appointments at the Universities of Sheffield, Edinburgh and Portsmouth, all in UK. Since 1999, he has been with Electronics and Computer Science, the

University of Southampton, UK, where he currently holds the post of Professor in Intelligent Systems and Signal Processing. Dr Chen's research interests include adaptive signal processing, wireless communications, modelling and identification of nonlinear systems, neural network and machine learning, intelligent control system design, evolutionary computation methods and optimisation. He has published over 550 research papers.

Dr. Chen is a Fellow of the United Kingdom Royal Academy of Engineering, a Fellow of IET, a Distinguished Adjunct Professor at King Abdulaziz University, Jeddah, Saudi Arabia, and an ISI highly cited researcher in engineering (March 2004).



Zesong Fei (S'03-M'07) received the Ph.D. degree in Electronic Engineering in 2004 from Beijing Institute of Technology (BIT), Beijing, China.

He is an Associate Professor in BIT and is currently with the Research Institute of Communication Technology (RICT) of BIT, where he is involved in the design of the next generation high-speed wireless communication. His research interests include mobile communication, channel coding and modulation, cognitive radio and cooperative networking.

He was chief investigator of China national Natural Science Fund project. He is the senior member of Chinese Institute of Electronics and China Institute of Communications.