Robust Joint Hybrid Transceiver Design for mmWave Full-Duplex MIMO Relay Systems

Yunlong Cai, Ying Xu, Qingjiang Shi, Benoit Champagne, and Lajos Hanzo

Abstract—The joint design of hybrid beamforming matrices is conceived for multiuser mmWave full-duplex (FD) multiple-input multiple-output (MIMO) relay-aided systems in the presence of realistic channel state information (CSI) errors. Specifically, considering a probabilistic CSI error model, we maximize the system's worst-case sum rate by jointly optimizing the base station's (BS) analog and digital beamforming matrices, plus the analog receive and transmit beamforming matrices of the relay station (RS) as well as its digital amplify-and-forward (AF) beamforming matrix under practical constraints. Explicitly, the transmit power constraints of the BS and RS, the residual self-interference power constraint of the RS, the per-user quality of service constraints and the unit-modulus constraints on the analog beamforming matrix elements are all taken into account. Since the resultant optimization problem is very challenging due to its highly nonlinear objective function and nonconvex coupling constraints, we first transform it into a more tractable form. We then develop a novel joint optimization algorithm based on the penalty dual decomposition (PDD) technique to solve the resultant problem. The proposed PDD-based algorithm performs double-loop iterations: the inner loop updates the optimization variables in a block coordinate descent fashion, while the outer loop adjusts the Lagrange multipliers and penalty parameter, hence ensuring convergence to the set of stationary solutions of the original problem. Our simulations show that mmWave FD hybrid MIMO relay systems relying on our new algorithm significantly outperform both their non-robust FD and conventional half-duplex counterparts.

Index Terms—Millimeter wave, full-duplex, multiuser MIMO relay, hybrid beamforming, robust transceiver design.

I. INTRODUCTION

Millimeter wave (mmWave) communications in the 30 - 300GHz band represent a salient next-generation technique for mitigating the spectrum shortage and supporting very high data rates [1]–[5]. However, owing to the excessive cost of radio frequency (RF) chains and analog-to-digital (A/D) converters at such high frequencies, the traditional fully digital transceiver architecture cannot be utilized in mmWave-based multiple antenna systems. As an alternative, a hybrid transceiver structure consisting of cascaded baseband digital and RF analog beamformers has been the focus of intense research. Indeed, this structure makes it possible to employ less RF chains than the number of antenna elements, while imposing a unity-modulus constraint on the elements of the analog beamforming matrix.

Some representative algorithms have been proposed in [6]–[10], [13] for the design of the hybrid transceiver parameters, in particular for the analog beamforming matrix. In [6], the authors have exploited the structure of mmWave channels to develop a hybrid transceiver algorithm based on orthogonal matching pursuit (OMP). A channel matching based hybrid beamforming algorithm has been proposed in [7]. To increase the performance of heuristic methods, the authors of [8] developed a joint hybrid transceiver design based on manifold optimization with antenna selection for large-scale multiple-input multiple-output (MIMO) mmWave systems. By taking the hardware constraint into account, a number of codebook-based hybrid beamforming techniques have been studied in [9]–[12]. In particular, an efficient hierarchical codebook design method has been developed in [11], [12] for beamforming training and channel estimation. Moreover, Full-duplex (FD) based mmWave systems have been investigated in [13].

Given the severe pathloss characterizing mmWave communications, relays may be invoked for avoiding LoS blockage or for extending the coverage area [14]. Hence, the study of relay-assisted mmWave communications is of great importance. Traditional relaying operates in half-duplex (HD) mode, which requires separate time or frequency resources for transmission and reception; by contrast, FD relaying potentially doubles the mmWave link capacity by supporting simultaneous transmission and reception [15]–[17]. However, mitigating the self-interference (SI) imposed by the high-power signal leakage from the transmitter output to the receiver input is a challenging issue in FD system design. Fortunately, major advances have been made in RF hardware for FD communications and hence, the SI can likely be mitigated to acceptable levels [18], [19]. Furthermore, several SI suppression algorithms have been developed for efficiently solving this problem in the digital domain, thereby motivating further research on transceiver design for FD relay-aided systems [20]–[23].

Recently, the authors of [24], [25] have designed

Y. Cai and Y. Xu are with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: ylcai@zju.edu.cn; xuying0124@zju.edu.cn).

Q. Shi is with the School of Software Engineering, Tongji University, Shanghai 201804, China (e-mail: qing.j.shi@gmail.com).

B. Champagne is with the Department of Electrical and Computer Engineering, McGill University, Montreal, QC, Canada, H3A 0E9 (e-mail: benoit.champagne@mcgill.ca).

L. Hanzo is with the Department of Electronics and Computer Science, University of Southampton, Southampton, UK (Email: lh@ecs.soton.ac.uk).

transceivers for HD mmWave MIMO relay systems, in order to improve network coverage. However, the existing HD mmWave relays rely on separate design approaches for the hybrid beamforming matrices. Indeed, the analog and digital beamforming matrices are not optimized jointly for maximizing the system's sum rate, which hence constitutes a heuristic, suboptimal approach from an optimization perspective. Finally, most of the mmWave MIMO relay designs stipulate the idealized simplifying assumption of having perfect channel state information (CSI), whilst there is a paucity of robust joint transceiver designs relying on realistic imperfect CSI.

To fill this gap, we jointly design the hybrid beamforming matrices of our sophisticated multiuser mmWave FD MIMO relay-aided system by assuming a probabilistic CSI error model. Based on this model, we first derive the worst case sum rate expression for our relaying system. Then, we jointly optimize the base station's (BS) analog and digital beamforming matrices, plus the relay station's (RS) analog transmit and receive beamforming matrices as well as its digital amplify-and-forward (AF) beamforming matrix by maximizing the worst-case sum rate under realistic optimization constraints. Explicitly, both the BS and RS transmit power constraints, the residual RS SI power constraint, as well as the per-user quality of service (QoS) constraints and the unity-modulus constraints of the RF analog beamforming matrix elements are taken into account.

However, this holistic global optimization problem cannot be solved directly due to its highly nonlinear objective function and nonconvex coupling constraints, imposed by the product of the digital and analog beamforming matrices as well as the unity-modulus constraint of the analog matrix elements. By introducing auxiliary variables, we first convert this worst-case sum-rate maximization problem into an equivalent but more tractable mean squared error minimization problem. Then, we invoke the innovative penalty dual decomposition (PDD) optimization method [26]–[29] for solving this simplified problem and jointly optimize the analog as well as digital beamforming matrices. The coupling components of the problem are handled by introducing new auxiliary variables and equality constraints. To elaborate, with the aid of the PDD method, we penalize and dualize the newly introduced equality constraints into the objective function as augmented Lagrangian (AL) components [30], [31]. The resultant augmented Lagrangian problem is then solved by our new double-loop iterative optimization algorithm. Explicitly, in the inner loop, we resort to the concave-convex procedure (CCCP) [32]–[34] for updating the optimization variables in a block coordinate descent fashion, where the subproblems of each block can be solved in closed form. Then in the outer loop, we adjust both the Lagrange multipliers and the penalty parameter of the AL cost function. The proposed robust PDD-based joint hybrid transceiver design algorithm converges to the set of stationary solutions of the original optimization problem. Furthermore, we extend the proposed algorithm to the practical scenario of finite-resolution phase shifters and analyze its computational complexity. Finally, we demonstrate that the proposed robust joint hybrid FD transceiver design significantly outperforms its nonrobust counterpart as well as

the conventional HD hybrid transceiver design.

The main original contributions of this work are summarized as follows:

1) To increase coverage and guarantee high transmission rates for mmWave communications, we propose a FD hybrid MIMO relay system architecture, and formulate a realistic but very challenging joint optimization problem for the optimization of the underlying transceiver parameter in the presence of CSI errors.
2) We transform this complex problem into an equivalent yet more tractable form and develop a novel PDD-based double-loop algorithm to solve it. The proposed robust algorithm converges to the set of stationary solutions of the original optimization problem.
3) The proposed algorithm is extended to the scenario of finite-resolution phase-shifter and a detailed computational complexity analysis is provided as well.
4) The properties of the new algorithm are investigated by means of simulations using a practical mmWave channel model; the results demonstrate its advantages over existing benchmarks schemes.

The rest of this paper is organized as follows. In Sections II and III, we present our system model and formulate the worst-case sum-rate maximization problem, respectively. In Section IV, we first transform the original problem into a more tractable but equivalent form and then develop the PDD-based joint hybrid transceiver design algorithm for solving the resultant problem. Our finite-resolution phase-shifter design and computational complexity analysis are discussed in Section V. In Section VI, we demonstrate the benefits of the proposed algorithms. Our conclusions are offered in Section VII, whilst proofs and background derivations appear in the Appendices.

Notations: $\mathbb{E}\{\cdot\}$ denotes the expectation operator. Boldface upper case and lower case letters represent matrices and vectors, respectively. $\text{Tr}(\mathbf{A})$, $\mathbf{A}^T$, and $\mathbf{A}^H$ represent the trace, transpose, and conjugate transpose of matrix $\mathbf{A}$, respectively. $\mathbf{A}(a:b,:)$ denotes a new matrix obtained by selecting rows a through b of matrix $\mathbf{A}$. For a matrix $\mathbf{A}$, $\mathbf{A}(i,j)$ denote the element at the intersection of row i and column j. $\mathbf{I}$ and $\mathbf{0}$ stand for identity and zero matrices with appropriate dimensions, respectively. $\| \cdot \|$ and $\det(\cdot)$ represent the Frobenius norm and determinant, respectively. $\Re\{.\}$, $\Im\{.\}$ and $|\cdot|$ respectively denote the real part, imaginary part and magnitude of a complex number. The operator $\text{vec}(\cdot)$ stacks the elements of a matrix in one long column vector, while $\text{diag}\{\cdot\}$ denotes a diagonal matrix with the given arguments on its diagonal. The symbol $\otimes$ denotes the Kronecker product. $\mathbb{C}^{m\times n}$ ($\mathbb{R}^{m\times n}$) denotes the space of $m \times n$ complex (real) matrices. The projection of a point $\mathbf{X}$ onto a set Ω is denoted by $\mathbb{P}_\Omega\{\mathbf{X}\} \triangleq \min_{\mathbf{Y}\in\Omega} \|\mathbf{X} - \mathbf{Y}\|$. If Ω is a sphere of radius r centered at the origin, i.e., $\Omega = \{\mathbf{X}|\|\mathbf{X}\| \leq r\}$, then $\mathbb{P}_\Omega\{\mathbf{X}\}$ is equal to $r\frac{\mathbf{X}}{\|\mathbf{X}\|+\max(0,r-\|\mathbf{X}\|)}$.

II. SYSTEM MODEL

Let us consider a multiuser mmWave FD MIMO relay system, as shown in Fig. 1, which consists of one BS, one FD RS and K users. The BS is equipped with N_t transmit antennas and R_1 ($R_1 \leq N_t$) RF chains. The RS, which operates in FD

mode, is equipped with N_r receive and N_r transmit antennas, as well as R_2 $(R_2 \leq N_r)$ RF chains for the receive and transmit processing, respectively. Each user is equipped with a single antenna. We assume that $K \leq \min\{R_1, R_2\}$ so as to provide sufficient degrees of freedom for signal detection. We consider a narrowband formulation with flat fading radio channel conditions. Furthermore, we assume that there is no direct link between the BS and the end users due to physical obstacles or severe attenuation.

Let $\mathbf{s} = [s_1, \ldots, s_K]^T$ represent the $K \times 1$ transmit symbol vector, whose elements are modeled as independent random variables with zero mean and unit variance, i.e., $\mathbb{E}\{|s_k|^2\} = 1$. The transmit signal vector at the BS can be expressed as

$$\mathbf{x}_B = \mathbf{VFs}, \tag{1}$$

where $\mathbf{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_K] \in \mathbb{C}^{R_1 \times K}$ denotes the BS digital beamforming matrix and $\mathbf{V} \in \mathbb{C}^{N_t \times R_1}$ denotes the BS analog transmit beamforming matrix. The transmit power of the BS can be expressed as

$$P_B \triangleq \mathbb{E}\{\text{Tr}(\mathbf{x}_B \mathbf{x}_B^H)\} = \|\mathbf{VF}\|^2. \tag{2}$$

The received signal at the RS can be written as

$$\mathbf{y}_R = \mathbf{GVFs} + \mathbf{n}_R + \mathbf{i}_R, \tag{3}$$

where $\mathbf{G} \in \mathbb{C}^{N_r \times N_t}$ denotes the channel matrix between the BS and the RS, $\mathbf{n}_R \in \mathbb{C}^{N_r \times 1}$ represents the complex circular Gaussian noise vector at the RS with zero mean and covariance $\mathbb{E}\{\mathbf{n}_R \mathbf{n}_R^H\} = \sigma_r^2 \mathbf{I}$, where σ_r^2 denotes the noise variance, and $\mathbf{i}_R$ represents the complex Gaussian residual SI with zero-mean and covariance $\mathbb{E}\{\mathbf{i}_R \mathbf{i}_R^H\} = \sigma_{si}^2 \mathbf{I}$. Here a constraint on the residual SI power is applied, i.e., $\mathbb{E}\{\|\mathbf{H}_R \mathbf{x}_R\|^2\} \leq P_{th}$ [23], where $\mathbf{H}_R$ denotes the relay residual SI channel[1], P_{th} denotes the maximum tolerable SI power, and $\mathbf{x}_R$ denotes the transmit signal at the RS. The latter is given by

$$\mathbf{x}_R = \mathbf{U}_2 \mathbf{W} \mathbf{U}_1 (\mathbf{GVFs} + \mathbf{n}_R + \mathbf{i}_R), \tag{4}$$

where $\mathbf{W} \in \mathbb{C}^{R_2 \times R_2}$ denotes the RS digital AF beamforming matrix, while $\mathbf{U}_1 \in \mathbb{C}^{R_2 \times N_r}$ and $\mathbf{U}_2 \in \mathbb{C}^{N_r \times R_2}$ represent the analog receive and transmit beamforming matrices at the RS, respectively. Defining $\sigma_R^2 \triangleq \sigma_r^2 + \sigma_{si}^2$, the transmit power at the RS can be expressed as

$$P_R \triangleq \mathbb{E}\{\text{Tr}(\mathbf{x}_R \mathbf{x}_R^H)\} = \|\mathbf{U}_2 \mathbf{W} \mathbf{U}_1 \mathbf{GVF}\|^2 + \sigma_R^2 \|\mathbf{U}_2 \mathbf{W} \mathbf{U}_1\|^2. \tag{5}$$

The received signal at user $k \in \mathcal{K} \triangleq \{1, \ldots, K\}$ can be written as

$$y_k = \sum_{j=1}^{K} \mathbf{h}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1 \mathbf{GVf}_j s_j + \mathbf{h}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1 (\mathbf{n}_R + \mathbf{i}_R) + n_k, \tag{6}$$

where $\mathbf{h}_k \in \mathbb{C}^{N_r \times 1}$ represents the Hermitian transpose of the channel vector between the RS and user k, while n_k is the complex circular Gaussian receiver noise at user k with zero mean and variance $\mathbb{E}\{|n_k|^2\} = \sigma_k^2$.

---

[1]The SI power can be initially mitigated to an acceptable level based on the use of hardware-aided FD interference cancellation techniques such as [16]–[19] and passive cancellation methods [20]. Consequently, herein the focus is on optimizing the hybrid beamforming algorithm to handle the residual SI.

The CSI from the BS to the RS and the CSI of the SI link at the RS can be obtained by implementing a suitable channel estimation algorithm, such as available from e.g., [35]–[37]. Due to the stationarity and high transmit power of the BS and RS, we assume that the corresponding channel matrices, i.e. $\mathbf{G}$ and $\mathbf{H}_R$, can be known with high accuracy. However, to obtain the CSI from the RS to the users, represented by vectors $\mathbf{h}_k$, is more challenging. Due to user mobility, processing latency and other limitations, the CSI for the user channels is inevitably corrupted by estimation errors. The channel matrix from the RS to the users, i.e., $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_K]^H$, can be expressed as

$$\mathbf{H} = \hat{\mathbf{H}} + \Delta\mathbf{H}, \tag{7}$$

where $\hat{\mathbf{H}} = [\hat{\mathbf{h}}_1, \ldots, \hat{\mathbf{h}}_K]^H$ denotes the estimated channel matrix and $\Delta\mathbf{H} = [\Delta\mathbf{h}_1, \ldots, \Delta\mathbf{h}_K]^H$ denotes the corresponding channel error matrix. Here, $\Delta\mathbf{H}$ is assumed to be statistically independent of the estimated channel matrix and characterized by a matrix-variate complex circular Gaussian distribution, i.e. $\text{vec}(\Delta\mathbf{H}) \sim \mathcal{CN}(\mathbf{0}, \sigma_e^2 \mathbf{I})$, where σ_e^2 denotes the variance of the individual CSI errors[2].

III. PROBLEM FORMULATION

In this section, we introduce the worst-case sum rate maximization problem under consideration in this work. Our starting point is the formulation of a deterministic lower bound on the sum rate. Let $\mathbf{p}_k \triangleq \mathbf{U}_2 \mathbf{W} \mathbf{U}_1 \mathbf{GVf}_k s_k$ denote the kth signal vector component transmitted by the RS (see eq. (6)) and let $\mathcal{H}(\mathbf{p}_k; y_k | \hat{\mathbf{H}})$ denote the conditional mutual information of user k, conditioned on estimated channel matrix $\hat{\mathbf{H}}$. Expanding $\mathcal{H}(\mathbf{p}_k; y_k | \hat{\mathbf{H}})$ in terms of the corresponding differential entropies yields,

$$\mathcal{H}(\mathbf{p}_k; y_k | \hat{\mathbf{H}}) = \mathcal{H}(\mathbf{p}_k | \hat{\mathbf{H}}) - \mathcal{H}(\mathbf{p}_k | y_k, \hat{\mathbf{H}}). \tag{8}$$

The first term on the right hand side (RHS) of (8) simplifies to $\frac{1}{2} \log \det(2\pi e \mathbf{Q}_k)$, where $\mathbf{Q}_k \triangleq \mathbb{E}\{\mathbf{p}_k \mathbf{p}_k^H\}$ denotes the transmit covariance matrix associated to $\mathbf{p}_k$. Regarding the second term, let us rewrite the received signal at user k as follows

$$\begin{aligned} y_k = &\hat{\mathbf{h}}_k^H \mathbf{p}_k + \sum_{k' \neq k}^{K} \hat{\mathbf{h}}_k^H \mathbf{p}_{k'} + \sum_{j=1}^{K} \Delta\mathbf{h}_k^H \mathbf{p}_j \\ &+ \hat{\mathbf{h}}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1 (\mathbf{n}_R + \mathbf{i}_R) \\ &+ \Delta\mathbf{h}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1 (\mathbf{n}_R + \mathbf{i}_R) + n_k. \end{aligned} \tag{9}$$

Consequently, the second term on the RHS of (8) is upper bounded by the entropy of a Gaussian random variable, i.e. [40],

$$\mathcal{H}(\mathbf{p}_k | y_k, \hat{\mathbf{H}}) \leq \frac{1}{2} \log \det\left(2\pi e \left(\mathbf{Q}_k - \frac{\mathbf{Q}_k \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H \mathbf{Q}_k}{\hat{\mathbf{h}}_k^H \mathbf{Q}_k \hat{\mathbf{h}}_k + \Upsilon_k}\right)\right) \tag{10}$$

---

[2]In practice, the mmWave CSI can be obtained by applying compressed sensing channel estimation algorithms, and under the case of additive Gaussian noise, the resulting channel estimation errors obey a multivariate Gaussian distribution [38], [39].
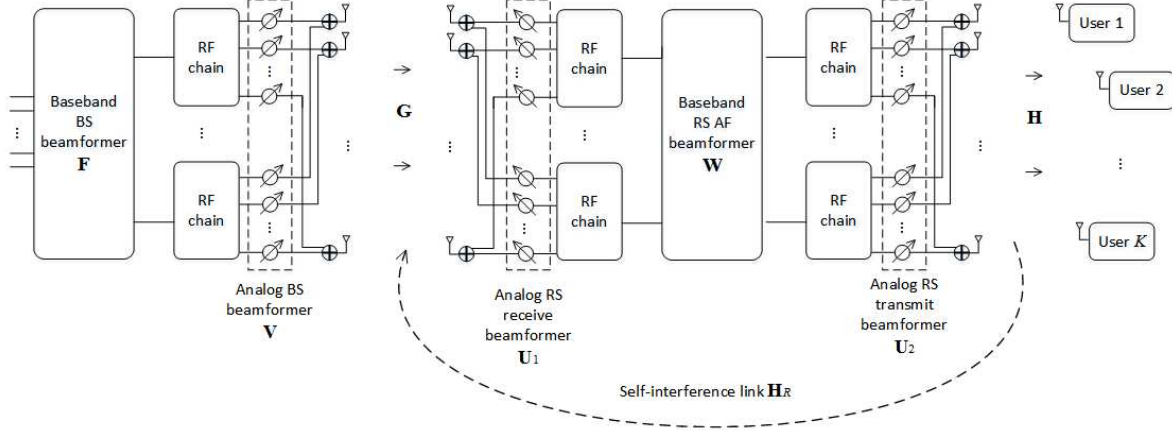
Fig. 1: Multiuser mmWave FD MIMO relay system

where

$$\Upsilon_k \triangleq \sum_{k' \neq k}^{K} \hat{\mathbf{h}}_k^H \mathbf{Q}_{k'} \hat{\mathbf{h}}_k + \sum_{j=1}^{K} \sigma_e^2 \mathrm{Tr}(\mathbf{Q}_j)$$
$$+ \sigma_R^2 \hat{\mathbf{h}}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1 \mathbf{U}_1^H \mathbf{W}^H \mathbf{U}_2^H \hat{\mathbf{h}}_k \qquad (11)$$
$$+ \sigma_R^2 \sigma_e^2 \mathrm{Tr}(\mathbf{U}_2 \mathbf{W} \mathbf{U}_1 \mathbf{U}_1^H \mathbf{W}^H \mathbf{U}_2^H) + \sigma_k^2.$$

Hence, substituting (10) into (8) and making use of the Woodbury matrix identity yields the following deterministic lower bound on the conditional mutual information for user k:

$$\mathcal{H}(\mathbf{p}_k; y_k | \hat{\mathbf{H}}) \geq \frac{1}{2} \log \left(1 + \frac{|\hat{\mathbf{h}}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1 \mathbf{G} \mathbf{f}_k|^2}{\Upsilon_k} \right). \quad (12)$$

The detailed derivation of (12) is presented in Appendix A.

Then, by considering the sum of the lower bound of the mutual information over all users, we can formulate the worst-case sum rate maximization optimization problem as shown in (13) where we define for convenience

$$N_k \triangleq \sigma_e^2 \left(\sum_{j=1}^{K} \|\mathbf{U}_2 \mathbf{W} \mathbf{U}_1 \mathbf{G} \mathbf{f}_j\|^2 + \sigma_R^2 \|\mathbf{U}_2 \mathbf{W} \mathbf{U}_1\|^2 \right) + \sigma_k^2.$$
$$(14)$$

The residual self-interference power constraint is shown in (13b) while constraints (13c) and (13d) reflect the bounded transmit power budget of the BS and RS, respectively. Constraint (13e) guarantees the per user QoS (i.e. transmission rate), where γ_k denotes the threshold of the achievable transmission rate for user k. Finally, the constant modulus constraint of the analog beamforming matrices is given by (13f).

The above problem formulation can be extended to more complex situations. For intance, when considering hardware impairments, e.g., limited analog-to-digital convertor accuracy, oscillator phase noise or low-noise-amplifier distortion, such sources of transmitter/receiver distortion in the communication chain can be modeled as independent Gaussian distributed errors [39]. In this case, we can still follow the same approach as introduced above to derive the corresponding lower bound on the system's sum rate.

We note that the constrained optimization problem (13) is very difficult to solve due to the highly coupled and nonconvex objective function and constraints. In the following section, we propose an efficient joint transceiver optimization design algorithm.

IV. ROBUST JOINT TRANSCEIVER DESIGN ALGORITHM

In this section, following a brief overview of the PDD optimization framework, we first transform the problem in (13) into a more tractable yet equivalent form. Then, by applying the PDD technique, we obtain an alternative formulation to (13) where a number of AL terms are incorporated into the objective in order to handle the exacting coupling constraints. We subsequently develop an efficient CCCP algorithm to solve the AL problem in the inner loop of the PDD algorithm in a block coordinate descent fashion. Finally, we summarize the proposed PDD-based joint transceiver design algorithm. The flow chart of our proposed approach is shown in Fig. 2.

A. Proposed PDD optimization framework

Below, we briefly review the PDD method in a general framework. Consider the following problem:

$$(P) \quad \min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x})$$
$$\text{s.t. } \mathbf{h}(\mathbf{x}) = \mathbf{0}, \qquad (15)$$
$$\mathbf{g}(\mathbf{x}) \leq \mathbf{0}.$$

where $f(\mathbf{x})$ is a scalar continuously differentiable function and $\mathcal{X} \subseteq \mathbb{R}^n$ is a closed convex set. As for constraints, $\mathbf{h}(\mathbf{x}) \in \mathbb{R}^p$ is a vector of p continuously differentiable functions, and $\mathbf{g}(\mathbf{x}) \in \mathbb{R}^q$ is a vector of q differentiable but possibly nonconvex functions. In order to handle the equality

$$\max_{\mathbf{F},\mathbf{W},\mathbf{V},\mathbf{U}_1,\mathbf{U}_2} \quad \sum_{k=1}^{K} \log\left(1 + \frac{|\hat{\mathbf{h}}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1 \mathbf{G} \mathbf{V} \mathbf{f}_k|^2}{\sum_{j\neq k}^{K} |\hat{\mathbf{h}}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1 \mathbf{G} \mathbf{V} \mathbf{f}_j|^2 + \sigma_R^2 \|\hat{\mathbf{h}}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1\|^2 + N_k}\right) \tag{13a}$$

$$\text{s.t.} \quad \|\mathbf{H}_R \mathbf{U}_2 \mathbf{W} \mathbf{U}_1 \mathbf{G} \mathbf{V} \mathbf{F}\|^2 + \sigma_R^2 \|\mathbf{H}_R \mathbf{U}_2 \mathbf{W} \mathbf{U}_1\|^2 \leq P_{th}, \tag{13b}$$

$$\|\mathbf{V}\mathbf{F}\|^2 \leq P_1, \tag{13c}$$

$$\|\mathbf{U}_2 \mathbf{W} \mathbf{U}_1 \mathbf{G} \mathbf{V} \mathbf{F}\|^2 + \sigma_R^2 \|\mathbf{U}_2 \mathbf{W} \mathbf{U}_1\|^2 \leq P_2, \tag{13d}$$

$$\log\left(1 + \frac{|\hat{\mathbf{h}}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1 \mathbf{G} \mathbf{V} \mathbf{f}_k|^2}{\sum_{j\neq k}^{K} |\hat{\mathbf{h}}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1 \mathbf{G} \mathbf{V} \mathbf{f}_j|^2 + \sigma_R^2 \|\hat{\mathbf{h}}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1\|^2 + N_k}\right) \geq \gamma_k, \forall k, \tag{13e}$$

$$|\mathbf{V}(i,j)| = 1, |\mathbf{U}_1(i,j)| = 1, |\mathbf{U}_2(i,j)| = 1, \forall i,j, \tag{13f}$$
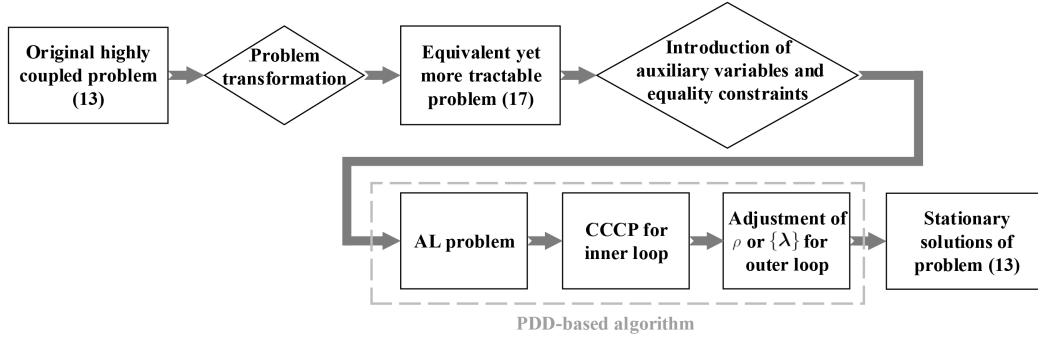

Fig. 2: The flow chart of the proposed PDD-based solution approach.

constraints, the PDD method aims for solving the following AL problem:

$$(P_{\varrho,\boldsymbol{\lambda}}) \quad \min_{\mathbf{x} \in \mathcal{X}} \mathcal{L}(\mathbf{x}) \triangleq f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x}) + \frac{1}{2\varrho}\|\mathbf{h}(\mathbf{x})\|^2$$
$$= f(\mathbf{x}) + \frac{1}{2\varrho}\|\mathbf{h}(\mathbf{x}) + \varrho\boldsymbol{\lambda}\|^2$$
$$\text{s.t. } \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \tag{16}$$

where $\mathcal{L}(\mathbf{x})$ is the AL function with scalar penalty parameter $\varrho > 0$ and dual variable $\boldsymbol{\lambda}$. In particular, when $\varrho \to \infty$, solving the above problem yields an identical solution to problem (15) [30]. The PDD method is a double-loop algorithm, where the inner loop solves the AL problem (16) via a block coordinate descent method while the outer loop updates the penalty parameter ϱ or the dual variable $\boldsymbol{\lambda}$ according to the constraint violation. The PDD algorithm is summarized in **Algorithm 1**, where the inner loop is subsumed by the "optimize" function in Step 2. A detailed discussion about the convergence and optimality properties of this algorithm can be found in [26], [27]. In particular, it can be shown that under appropriate conditions, the sequence of iterates $\mathbf{x}^r$ generated by the PDD method converges to a KKT (stationary) point of the original

problem[3].

Algorithm 1 PDD algorithm for problem (15)

0. initialize $\mathbf{x}^0 \in \mathcal{X}$, $\varrho_0 > 0$, $\boldsymbol{\lambda}_0$, and set $0 < c < 1$, $r = 0$
1. **Repeat**
2. $\quad \mathbf{x}^{r+1} = \texttt{optimize}(P_{\varrho_r,\boldsymbol{\lambda}_r}, \mathbf{x}^r)$ % Solve AL problem
3. $\quad$ **if** $\|\mathbf{h}(\mathbf{x}^r)\|_\infty \leq \eta_r$
4. $\quad\quad\quad \boldsymbol{\lambda}_{r+1} = \boldsymbol{\lambda}_r + \frac{1}{\varrho_r}\mathbf{h}(\mathbf{x}^r)$
5. $\quad\quad\quad \varrho_{r+1} = \varrho_r$
6. $\quad$ **else**
7. $\quad\quad\quad \boldsymbol{\lambda}_{r+1} = \boldsymbol{\lambda}_r$
8. $\quad\quad\quad \varrho_{r+1} = c\varrho_r$
9. $\quad$ **end**
10. $\quad r = r + 1$
11. **Until** some termination criterion is met

---

[3]In the absence of better alternative, it is readily seen that this proposed algorithm is the best choice for solving this optimization problem under study. Due to the NP-hard nature of the problem, based on existing optimization techniques, it does not seem possible to provide a quantitative analysis of the performance gap between the optimal solution and that obtained with the proposed algorithm. At the present time, ensuring convergence to a stationary point is the best we can do in terms of convergence analysis for this kind of constrained optimization problem. The characterization of the performance gap between the optimal and iterative solutions remains an open problem for future research.

B. Problem transformation

We first transform the optimization problem in (13) into an equivalent yet more tractable form. To this end, let us introduce the auxiliary variables $u_k, z_k > 0$ ($k \in \mathcal{K}$), which stand for the receiver gain and weighting coefficient of user k, respectively. Using these variables, (13) can be equivalently formulated as the following weighted MSE minimization problem shown in (17), in the sense that both problems share the same global optimal solutions for $\mathbf{F}$, $\mathbf{W}$, $\mathbf{V}$, $\mathbf{U}_1$ and $\mathbf{U}_2$ under the given constraints, where $\gamma_k^{'} \triangleq 2^{\gamma_k}$ and e_k denotes the MSE of user k, which can be expressed as

$$
\begin{aligned}
e_k &\triangleq \mathbb{E}\left[|u_k y_k - s_k|^2\right] \\
&= |u_k|^2 \left(\sum_{j=1}^{K} |\hat{\mathbf{h}}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1 \mathbf{G} \mathbf{V} \mathbf{f}_j|^2 + \sigma_R^2 \|\hat{\mathbf{h}}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1\|^2 + N_k \right) \\
&\quad - \left(u_k \hat{\mathbf{h}}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1 \mathbf{G} \mathbf{V} \mathbf{f}_k + u_k^* \mathbf{f}_k^H \mathbf{V}^H \mathbf{G}^H \mathbf{U}_1^H \mathbf{W}^H \mathbf{U}_2^H \hat{\mathbf{h}}_k \right) + 1.
\end{aligned}
\tag{18}
$$

The detailed proof of the equivalence between (13) and (17) can be found in Appendix B.

C. Augmented Lagrangian problem

Within the PDD optimization framework the difficulties posed by the coupling terms in the constraints (13b)-(13e) can be handled by introducing auxiliary variables and additional equality constraints. In this way, the original coupling constraints can be effectively moved into the objective function of an AL problem, such that the search variables can be optimized in a block coordinate descent manner [41]. While auxiliary variables and equality constraints can be introduced in a different way, here we make use of the following guidelines, which based on our experience with the PDD, help to simplify the solution of the resulting AL problem:

1) The new constraints should not contain coupling terms;
2) It should be possible to jointly optimize (i.e., within one step of block coordinate descent) the variables appearing in the same constraint;
3) Each search variable cannot appear in more than one constraint.

Essentially, rule 1) guarantees that the resulting algorithm converges to a stationary solution of the original problem; we note that conventional alternating optimization methods with coupling constraints cannot guarantee optimality [41]. Rule 2) prevents that the updating algorithm gets trapped in a deadlock while updating search variables. Rule 3) makes it possible to decompose the problem into a number of subproblems, which can be easily solved either in a closed form or by the Lagrange multiplier method.

In order to cope with the coupling constraints attached to problem (17), let us introduce a set of auxiliary variables and equality constraints. Based on rules 1) and 2), we introduce auxiliary variables: $\mathbf{S}$, $\mathbf{Y}$ and $\tilde{\mathbf{X}}$, with equality constraints $\mathbf{S} = \mathbf{V} \mathbf{F}$, $\mathbf{Y} = \mathbf{U}_2 \mathbf{W} \mathbf{U}_1$ and $\tilde{\mathbf{X}} = \mathbf{Y} \mathbf{G} \mathbf{S}$. According to rule 3), we introduce auxiliary variables: $\mathbf{X}$, $\hat{\mathbf{Y}}$, $\tilde{\mathbf{S}}$, $\bar{\mathbf{Y}}$, $\bar{\mathbf{X}}$ and $\{\hat{\mathbf{X}}_k, \hat{\mathbf{Y}}_k\}$, with constraints $\mathbf{X} = \tilde{\mathbf{X}}$, $\hat{\mathbf{Y}} = \sigma_R \mathbf{Y}$, $\tilde{\mathbf{S}} = \mathbf{S}$, $\bar{\mathbf{Y}} = \sigma_R \mathbf{H}_R \mathbf{Y}$, $\bar{\mathbf{X}} = \mathbf{H}_R \mathbf{X}$,

$\hat{\mathbf{X}}_k = \mathbf{X}$, and $\hat{\mathbf{Y}}_k = \mathbf{Y}$ ($k \in \mathcal{K}$). Problem (17) can then be equivalently expressed as (19), where $\mathcal{Z} \triangleq \{u_k, z_k, \mathbf{F}, \mathbf{W}, \mathbf{V}, \mathbf{U}_1, \mathbf{U}_2, \mathbf{S}, \mathbf{Y}, \mathbf{X}, \tilde{\mathbf{X}}, \hat{\mathbf{Y}}, \tilde{\mathbf{S}}, \bar{\mathbf{Y}}, \bar{\mathbf{X}}, \hat{\mathbf{X}}_k, \hat{\mathbf{Y}}_k\}$ denotes the complete set of variables, $\mathbf{a}_k = [\underbrace{0, \ldots, 0}_{k-1}, 1, \underbrace{0, \ldots, 0}_{K-k}]^T$ (with the 1 in the kth position) is used as a selection vector, and

$$
\begin{aligned}
\tilde{e}_k &\triangleq |u_k|^2 \Big(\sum_{j=1}^{K} |\hat{\mathbf{h}}_k^H \mathbf{X} \mathbf{a}_j|^2 + \sigma_R^2 \|\hat{\mathbf{h}}_k^H \mathbf{Y}\|^2 \\
&\quad + \sigma_e^2 \left(\|\mathbf{X}\|^2 + \sigma_R^2 \|\mathbf{Y}\|^2 \right) + \sigma_k^2 \Big) \\
&\quad - \left(u_k \hat{\mathbf{h}}_k^H \mathbf{X} \mathbf{a}_k + u_k^* \mathbf{a}_k^T \mathbf{X}^H \hat{\mathbf{h}}_k \right) + 1.
\end{aligned}
\tag{20}
$$

Furthermore, we note that constraint (19e) can be viewed as a difference of convex (DC) functions:

$$
\begin{aligned}
&\sum_{j \neq k}^{K} |\hat{\mathbf{h}}_k^H \hat{\mathbf{X}}_k \mathbf{a}_j|^2 + \sigma_R^2 \|\hat{\mathbf{h}}_k^H \hat{\mathbf{Y}}_k\|^2 + \sigma_e^2 \Big(\sum_{j=1}^{K} \|\hat{\mathbf{X}}_k \mathbf{a}_j\|^2 + \sigma_R^2 \|\hat{\mathbf{Y}}_k\|^2 \Big) \\
&\quad + \sigma_k^2 - \frac{1}{\gamma_k^{'}} |\hat{\mathbf{h}}_k^H \hat{\mathbf{X}}_k \mathbf{a}_k|^2 \leq 0, \; \forall k,
\end{aligned}
\tag{21}
$$

which we seek to transform into a convex constraint. Based on the CCCP concept [32], [33], by linearizing the term $|\hat{\mathbf{h}}_k^H \hat{\mathbf{X}}_k \mathbf{a}_k|^2$ in the ith iteration around the current point $\hat{\mathbf{X}}_k^{(i)}$, the following approximated convex constraint is obtained:

$$
\begin{aligned}
&\sum_{j \neq k}^{K} |\hat{\mathbf{h}}_k^H \hat{\mathbf{X}}_k \mathbf{a}_j|^2 + \sigma_R^2 \|\hat{\mathbf{h}}_k^H \hat{\mathbf{Y}}_k\|^2 + \sigma_e^2 \Big(\sum_{j=1}^{K} \|\hat{\mathbf{X}}_k \mathbf{a}_j\|^2 + \sigma_R^2 \|\hat{\mathbf{Y}}_k\|^2 \Big) \\
&\quad + \sigma_k^2 + \frac{1}{\gamma_k^{'}} |\hat{\mathbf{h}}_k^H \hat{\mathbf{X}}_k^{(i)} \mathbf{a}_k|^2 \\
&\quad - \frac{2}{\gamma_k^{'}} \Re\{\mathbf{a}_k^H \hat{\mathbf{X}}_k^{(i)H} \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H \hat{\mathbf{X}}_k \mathbf{a}_k\} \leq 0, \; \forall k.
\end{aligned}
\tag{22}
$$

We then take all the equality constraints into account by augmenting the objective function (19a) with Lagrange multipliers: $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3, \boldsymbol{\lambda}_4, \boldsymbol{\lambda}_5, \boldsymbol{\lambda}_6, \boldsymbol{\lambda}_7, \boldsymbol{\lambda}_8, \{\boldsymbol{\lambda}_{9,k}\}, \{\boldsymbol{\lambda}_{10,k}\}$, and a penalty coefficient ρ. Therefore, by employing the PDD method discussed in Subsection IV-A, the AL problem can be formulated as (23). The proposed PDD-based transceiver design algorithm exhibits a double-loop structure, where the outer loop updates the dual variables and the penalty parameter while the inner loop seeks to optimize the primal variables by solving problem (23).

D. Proposed CCCP algorithm for solving problem (23)

Referring to our discussion in Subsection IV-A, the central element of the PDD-based algorithm is the inner loop for solving the AL problem, represented by the `optimize` function in the outer loop of **Algorithm 1**. For the FD MIMO relay transceiver design under consideration here, the core AL problem takes the form of (23), as obtained after the introduction of auxiliary variables and application of the CCCP. In the following, we focus on solving the corresponding CCCP problem (23) in a block coordinate descent fashion, in which we divide the optimization variables into a number of

$$\min_{\mathbf{F},\mathbf{W},\mathbf{U}_1,\mathbf{U}_2,\mathbf{V},\{u_k,z_k\}} \quad \sum_{k=1}^{K} z_k e_k - \log(z_k)$$
$$\text{s.t.} \quad (13b) - (13d), (13f) \tag{17}$$
$$\frac{|\hat{\mathbf{h}}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1 \mathbf{G} \mathbf{V} \mathbf{f}_k|^2}{\sum_{j\neq k}^{K} |\hat{\mathbf{h}}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1 \mathbf{G} \mathbf{V} \mathbf{f}_j|^2 + \sigma_R^2 \|\hat{\mathbf{h}}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1\|^2 + N_k} \geq \gamma_k',$$

$$\min_{\mathcal{Z}} \sum_{k=1}^{K} z_k \tilde{e}_k - \log(z_k) \tag{19a}$$
$$\text{s.t.} \quad \|\bar{\mathbf{X}}\|^2 + \|\bar{\mathbf{Y}}\|^2 \leq P_{th}, \tag{19b}$$
$$\|\tilde{\mathbf{S}}\|^2 \leq P_1, \tag{19c}$$
$$\|\tilde{\mathbf{X}}\|^2 + \|\tilde{\mathbf{Y}}\|^2 \leq P_2, \tag{19d}$$
$$\frac{|\hat{\mathbf{h}}_k^H \hat{\mathbf{X}}_k \mathbf{a}_k|^2}{\sum_{j\neq k}^{K} |\hat{\mathbf{h}}_k^H \hat{\mathbf{X}}_k \mathbf{a}_j|^2 + \sigma_R^2 |\hat{\mathbf{h}}_k^H \hat{\mathbf{Y}}_k|^2 + \sigma_e^2 (\sum_{j=1}^{K} \|\hat{\mathbf{X}}_k \mathbf{a}_j\|^2 + \sigma_R^2 \|\hat{\mathbf{Y}}_k\|^2) + \sigma_k^2} \geq \gamma_k', \forall k, \tag{19e}$$
$$\mathbf{S} = \mathbf{VF}, \ \mathbf{Y} = \mathbf{U}_2 \mathbf{W} \mathbf{U}_1, \ \tilde{\mathbf{X}} = \mathbf{YGS}, \ \tilde{\mathbf{X}} = \mathbf{X}, \ \tilde{\mathbf{Y}} = \sigma_R \mathbf{Y}, \ \tilde{\mathbf{S}} = \mathbf{S}, \tag{19f}$$
$$\bar{\mathbf{Y}} = \sigma_R \mathbf{H}_R \mathbf{Y}, \ \bar{\mathbf{X}} = \mathbf{H}_R \mathbf{X}, \ \hat{\mathbf{X}}_k = \mathbf{X}, \ \hat{\mathbf{Y}}_k = \mathbf{Y}, \forall k, \tag{19g}$$
$$|\mathbf{V}(i,j)| = 1, |\mathbf{U}_1(i,j)| = 1, |\mathbf{U}_2(i,j)| = 1, \forall i,j \tag{19h}$$

$$\begin{aligned}
\min_{\mathcal{Z}} \quad & \sum_{k=1}^{K} (z_k \tilde{e}_k - \log(z_k)) + \frac{1}{2\rho} \big(\|\mathbf{U}_2 \mathbf{W} \mathbf{U}_1 - \mathbf{Y} + \rho\boldsymbol{\lambda}_1\|^2 + \|\mathbf{VF} - \mathbf{S} + \rho\boldsymbol{\lambda}_2\|^2 \\
& + \|\mathbf{YGS} - \tilde{\mathbf{X}} + \rho\boldsymbol{\lambda}_3\|^2 + \|\mathbf{H}_R \mathbf{X} - \bar{\mathbf{X}} + \rho\boldsymbol{\lambda}_4\|^2 + \|\mathbf{X} - \tilde{\mathbf{X}} + \rho\boldsymbol{\lambda}_5\|^2 \\
& + \|\sigma_R \mathbf{H}_R \mathbf{Y} - \bar{\mathbf{Y}} + \rho\boldsymbol{\lambda}_6\|^2 + \|\sigma_R \mathbf{Y} - \tilde{\mathbf{Y}} + \rho\boldsymbol{\lambda}_7\|^2 + \|\tilde{\mathbf{S}} - \mathbf{S} + \rho\boldsymbol{\lambda}_8\|^2 \\
& + \sum_{k=1}^{K} (\|\hat{\mathbf{X}}_k - \mathbf{X} + \rho\boldsymbol{\lambda}_{9,k}\|^2 + \|\hat{\mathbf{Y}}_k - \mathbf{Y} + \rho\boldsymbol{\lambda}_{10,k}\|^2) \big)
\end{aligned} \tag{23}$$
$$\text{s.t.} \quad (19b) - (19d), (22), (19h).$$

blocks, such that for each block the corresponding subproblem can be solved efficiently in closed form. Specifically, we partition the search variables into four blocks, and for each block, obtain the solution: the corresponding developments are presented under Steps 1 to 4 below, respectively.

In **Step 1**, we optimize $\{u_k\}, (\bar{\mathbf{X}}, \bar{\mathbf{Y}}), \tilde{\mathbf{S}}, \mathbf{V}, \mathbf{U}_1, \{\hat{\mathbf{X}}_k, \hat{\mathbf{Y}}_k\}$ in parallel by fixing the other variables. Note that in this case problem (23) can be decomposed into six independent subproblems.

The first subproblem for u_k is an unconstrained quadratic optimization problem. By examining the first order optimality condition, we can obtain the solution shown in (24).

The second subproblem with respect to $(\bar{\mathbf{X}}, \bar{\mathbf{Y}})$ is given by

$$\min_{\bar{\mathbf{Y}}, \bar{\mathbf{X}}} \quad \|\mathbf{H}_R \mathbf{X} - \bar{\mathbf{X}} + \rho\boldsymbol{\lambda}_4\|^2 + \|\sigma_R \mathbf{H}_R \mathbf{Y} - \bar{\mathbf{Y}} + \rho\boldsymbol{\lambda}_6\|^2$$
$$\text{s.t.} \quad \|\bar{\mathbf{X}}\|^2 + \|\bar{\mathbf{Y}}\|^2 \leq P_{th}. \tag{25}$$

It is readily seen that (25) is equivalent to the projection of a point onto a sphere centered at the origin, which admits a

closed-form solution as follows

$$(\bar{\mathbf{X}}, \bar{\mathbf{Y}}) = \mathbb{P}_{\Omega_1} \{\rho\boldsymbol{\lambda}_4 + \mathbf{H}_R \mathbf{X}, \rho\boldsymbol{\lambda}_6 + \sigma_R \mathbf{H}_R \mathbf{Y}\}, \tag{26}$$

where $\Omega_1 \triangleq \|\bar{\mathbf{X}}\|^2 + \|\bar{\mathbf{Y}}\|^2 \leq P_{th}$.

Similarly, the third subproblem with respect to $\tilde{\mathbf{S}}$ can be solved with the aid of projection, with solution given by

$$\tilde{\mathbf{S}} = \mathbb{P}_{\Omega_2} \{\mathbf{S} - \rho\boldsymbol{\lambda}_8\}, \tag{27}$$

where $\Omega_2 \triangleq \{\tilde{\mathbf{S}} | \|\tilde{\mathbf{S}}\|^2 \leq P_1\}$.

The fourth subproblem with respect to $\mathbf{V}$ is given by

$$\min_{|\mathbf{V}(i,j)|=1} \|\mathbf{VF} - \mathbf{S} + \rho\boldsymbol{\lambda}_2\|^2. \tag{28}$$

By appropriate rearrangement, it can be equivalently formulated as

$$\min_{|\mathbf{V}(i,j)|=1} \text{Tr}(\mathbf{V}^H \mathbf{V} \mathbf{C}_V) - 2\Re\{\text{Tr}(\mathbf{V}^H \mathbf{B}_V)\} \tag{29}$$

where $\mathbf{C}_V \triangleq \mathbf{FF}^H$ and $\mathbf{B}_V \triangleq (\mathbf{S} - \rho\boldsymbol{\lambda}_2)\mathbf{F}^H$. Since the unit modulus constraints are separable, we can use the one-iteration block coordinate descent type algorithm presented in the Appendix of [28] to solve problem (29).

$$u_k = \frac{\mathbf{x}_k^H \hat{\mathbf{h}}_k}{\sum_{j=1}^{K} |\hat{\mathbf{h}}_k^H \mathbf{x}_j|^2 + \sigma_R^2 \|\hat{\mathbf{h}}_k^H \mathbf{Y}\|^2 + \sigma_e^2 (\|\mathbf{X}\|^2 + \sigma_R^2 \|\mathbf{Y}\|^2) + \sigma_k^2}, \quad \forall k. \tag{24}$$

Similarly, the fifth subproblem with respect to $\mathbf{U}_1$ is given by

$$\min_{|\mathbf{U}_1(i,j)|=1} \|\mathbf{U}_2 \mathbf{W} \mathbf{U}_1 - \mathbf{Y} + \rho \boldsymbol{\lambda}_1\|^2. \tag{30}$$

Proceeding as in (28), (29), it can be equivalently formulated as

$$\min_{|\mathbf{U}_1(i,j)|=1} \mathrm{Tr}(\mathbf{U}_1^H \mathbf{A}_{U_1} \mathbf{U}_1) - 2\Re\{\mathrm{Tr}(\mathbf{U}_1^H \mathbf{B}_{U_1})\}, \tag{31}$$

where $\mathbf{A}_{U_1} \triangleq \mathbf{W}^H \mathbf{U}_2^H \mathbf{U}_2 \mathbf{W}$ and $\mathbf{B}_{U_1} \triangleq \mathbf{W}^H \mathbf{U}_2^H (\mathbf{Y} - \rho \boldsymbol{\lambda}_1)$. Problem (31) can also be solved based on the method in [28].

The last subproblem with regard to $\{\hat{\mathbf{X}}_k, \hat{\mathbf{Y}}_k\}, \forall k$, is expressed as

$$\min_{\{\hat{\mathbf{X}}_k, \hat{\mathbf{Y}}_k\}} \sum_{k=1}^{K} (\|\hat{\mathbf{X}}_k - \mathbf{X} + \rho \boldsymbol{\lambda}_{9,k}\|^2 + \|\hat{\mathbf{Y}}_k - \mathbf{Y} + \rho \boldsymbol{\lambda}_{10,k}\|^2))$$
$$\text{s.t.} \quad (22). \tag{32}$$

We can solve it by applying the method of Lagrange multiplier. The detailed derivation is shown in Appendix C.

In **Step 2**, we optimize $(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}})$, $\{z_k\}$, $\mathbf{U}_2$ and $\mathbf{F}$ in parallel while fixing the other variables. In this case, problem (23) can be decomposed into four independent unconstrained subproblems.

The first subproblem with respect to $(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}})$ can be solved with the aid of a projection as discussed before, and the solution is given by

$$(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}) = \mathbb{P}_{\Omega_3} \{ \frac{1}{2} (\rho \boldsymbol{\lambda}_5 + \mathbf{X} + \rho \boldsymbol{\lambda}_3 + \mathbf{Y} \mathbf{G} \mathbf{S}), \rho \boldsymbol{\lambda}_7 + \sigma_R \mathbf{Y} \}, \tag{33}$$

where $\Omega_3 \triangleq \{(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}) | \|\tilde{\mathbf{X}}\|^2 + \|\tilde{\mathbf{Y}}\|^2 \le P_2\}$.

The second subproblem with respect to z_k is given by

$$\min_{z_k} \quad z_k \tilde{e}_k - \log(z_k). \tag{34}$$

By examining the first order optimality condition for z_k, we obtain

$$z_k = \frac{1}{\tilde{e}_k}. \tag{35}$$

The third subproblem with respect to $\mathbf{U}_2$ is a quadratic optimization problem with unit modulus constraints, which is given by

$$\min_{|\mathbf{U}_2(i,j)|=1} \|\mathbf{U}_2 \mathbf{W} \mathbf{U}_1 - \mathbf{Y} + \rho \boldsymbol{\lambda}_1\|^2. \tag{36}$$

We can adopt the same method as applied to (29) to solve this problem.

The fourth subproblem of $\mathbf{F}$ can be expressed as

$$\min_{\mathbf{F}} \quad \|\mathbf{V} \mathbf{F} - \mathbf{S} + \rho \boldsymbol{\lambda}_2\|^2. \tag{37}$$

By applying the first order optimality condition, we obtain the unique solution for $\mathbf{F}$, which is given by

$$\mathbf{F} = (\mathbf{V}^H \mathbf{V})^{-1} \mathbf{V}^H (\mathbf{S} - \rho \boldsymbol{\lambda}_2). \tag{38}$$

In **Step 3**, we optimize $\mathbf{X}$ and $\mathbf{Y}$ in parallel by fixing the remaining variables. In this case, problem (23) can be decomposed into two independent unconstrained subproblems.

The subproblem with respect to $\mathbf{X}$ is given in (39). In this case, the first order optimality condition yields

$$\mathbf{X} = (\sum_{k=1}^{K} z_k |u_k|^2 (\hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H + \sigma_e^2 \mathbf{I}) + \frac{1}{2\rho} (\mathbf{H}_R^H \mathbf{H}_R + (K+1)\mathbf{I}))^{-1} \mathbf{M}, \tag{40}$$

where $\mathbf{M} \triangleq \sum_{k=1}^{K} z_k u_k^* \hat{\mathbf{h}}_k \mathbf{a}_k^T + \frac{1}{2\rho} \left(\mathbf{H}_R^H (\bar{\mathbf{X}} - \rho \boldsymbol{\lambda}_4) + \tilde{\mathbf{X}} - \rho \boldsymbol{\lambda}_5 + \sum_{k=1}^{K} (\rho \boldsymbol{\lambda}_{9,k} + \hat{\mathbf{X}}_k) \right)$.

The subproblem with respect to $\mathbf{Y}$ is given in (41). By vectorizing $\mathbf{Y}$ and applying the first order optimality condition, we obtain

$$\mathrm{vec}(\mathbf{Y}) = \mathbf{P}^{-1} \mathbf{q} \tag{42}$$

where

$$\mathbf{P} = 2\rho \sum_{k=1}^{K} z_k |u_k|^2 \sigma_R^2 \left(\mathbf{I} \otimes \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H + \sigma_e^2 \mathbf{I} \right) + (\sigma_R^2 + K + 1)\mathbf{I} + \sigma_R^2 \mathbf{I} \otimes \mathbf{H}_R^H \mathbf{H}_R + (\mathbf{G} \mathbf{S} \mathbf{S}^H \mathbf{G}^H)^T \otimes \mathbf{I}, \tag{43}$$

$$\mathbf{q} = \mathrm{vec}(\mathbf{U}_2 \mathbf{W} \mathbf{U}_1 + \rho \boldsymbol{\lambda}_1) + \mathrm{vec}\left((\tilde{\mathbf{X}} - \rho \boldsymbol{\lambda}_3) \mathbf{S}^H \mathbf{G}^H \right)$$
$$+ \sigma_R \mathrm{vec}\left(\mathbf{H}_R^H (\bar{\mathbf{Y}} - \rho \boldsymbol{\lambda}_6) \right) + \sigma_R \mathrm{vec}\left(\tilde{\mathbf{Y}} - \rho \boldsymbol{\lambda}_7 \right)$$
$$+ \sum_{k=1}^{K} \mathrm{vec}\left(\rho \boldsymbol{\lambda}_{10,k} + \hat{\mathbf{Y}}_k \right). \tag{44}$$

In **Step 4**, we optimize $\mathbf{S}$ and $\mathbf{W}$ in parallel by fixing the remaining variables.

The corresponding unconstrained subproblem for $\mathbf{S}$ is given by

$$\min_{\mathbf{S}} \quad \|\mathbf{V} \mathbf{F} - \mathbf{S} + \rho \boldsymbol{\lambda}_2\|^2 + \|\mathbf{Y} \mathbf{G} \mathbf{S} - \tilde{\mathbf{X}} + \rho \boldsymbol{\lambda}_3\|^2 + \|\tilde{\mathbf{S}} - \mathbf{S} + \rho \boldsymbol{\lambda}_8\|^2. \tag{45}$$

By examining the first order optimality condition, we obtain the optimal value of $\mathbf{S}$ as

$$\mathbf{S} = \left(2\mathbf{I} + \mathbf{G}^H \mathbf{Y}^H \mathbf{Y} \mathbf{G} \right)^{-1}$$
$$\times \left(\rho \boldsymbol{\lambda}_2 + \mathbf{V} \mathbf{F} + \mathbf{G}^H \mathbf{Y}^H (\tilde{\mathbf{X}} - \rho \boldsymbol{\lambda}_3) + \tilde{\mathbf{S}} + \rho \boldsymbol{\lambda}_8 \right). \tag{46}$$

The second subproblem with regard to $\mathbf{W}$ is given by

$$\min_{\mathbf{W}} \quad \|\mathbf{U}_2 \mathbf{W} \mathbf{U}_1 - \mathbf{Y} + \rho \boldsymbol{\lambda}_1\|^2. \tag{47}$$

By vectorizing $\mathbf{W}$ and examining the first order optimality condition, we obtain the following solution

$$\mathrm{vec}(\mathbf{W}) = \left((\mathbf{U}_1 \mathbf{U}_1^H)^T \otimes \mathbf{U}_2^H \mathbf{U}_2 \right)^{-1} \mathrm{vec}\left(\mathbf{U}_2^H (\mathbf{Y} - \rho \boldsymbol{\lambda}_1) \mathbf{U}_1^H \right). \tag{48}$$

$$\min_{\mathbf{X}} \quad \sum_{k=1}^{K} z_k \left(|u_k|^2 (\|\hat{\mathbf{h}}_k^H \mathbf{X}\|^2 + \sigma_e^2 \|\mathbf{X}\|^2) - (u_k \hat{\mathbf{h}}_k^H \mathbf{X} \mathbf{a}_k + u_k^* \mathbf{a}_k^T \mathbf{X}^H \hat{\mathbf{h}}_k) \right)$$

$$+ \frac{1}{2\rho} \left(\|\mathbf{H}_R \mathbf{X} - \bar{\mathbf{X}} + \rho \boldsymbol{\lambda}_4\|^2 + \|\mathbf{X} - \tilde{\mathbf{X}} + \rho \boldsymbol{\lambda}_5\|^2 + \sum_{k=1}^{K} (\|\hat{\mathbf{X}}_k - \mathbf{X} + \rho \boldsymbol{\lambda}_{9,k}\|^2) \right). \tag{39}$$

$$\min_{\mathbf{Y}} \quad \sum_{k=1}^{K} z_k |u_k|^2 \sigma_R^2 \left(\|\hat{\mathbf{h}}_k^H \mathbf{Y}\|^2 + \sigma_e^2 \|\mathbf{Y}\|^2 \right) + \frac{1}{2\rho} \left(\|\mathbf{U}_2 \mathbf{W} \mathbf{U}_1 - \mathbf{Y} + \rho \boldsymbol{\lambda}_1\|^2 \right.$$

$$+ \|\mathbf{Y}\mathbf{G}\mathbf{S} - \tilde{\mathbf{X}} + \rho \boldsymbol{\lambda}_3\|^2 + \|\sigma_R \mathbf{H}_R \mathbf{Y} - \bar{\mathbf{Y}} + \rho \boldsymbol{\lambda}_6\|^2 + \|\sigma_R \mathbf{Y} - \tilde{\mathbf{Y}} + \rho \boldsymbol{\lambda}_7\|^2 \tag{41}$$

$$\left. + \sum_{k=1}^{K} \|\hat{\mathbf{Y}}_k - \mathbf{Y} + \rho \boldsymbol{\lambda}_{10,k}\|^2 \right).$$

In each iteration of the the inner loop of the PDD-based algorithm (i.e., Step 2 in **Algorithm 1**), we implement the above four updating steps according to the block structure of the optimization variables. The resulting algorithm to solve (23) is summarized under the name of CCCP-based algorithm in **Algorithm 2**.

Algorithm 2 Proposed CCCP algorithm in the inner loop

---

0. Define the tolerance of accuracy ϵ_1 and the maximum number of iterations $N_{\max}$. Initialize the algorithm with a feasible point. Set the iteration number $i = 0$.
1. **Repeat**
2. – Update $\{u_k\}$, $(\bar{\mathbf{X}}, \bar{\mathbf{Y}})$, $\tilde{\mathbf{S}}$, $\mathbf{V}$, $\mathbf{U}_1$, $\{\hat{\mathbf{X}}_k, \hat{\mathbf{Y}}_k\}$ in **Step 1**.
3. – Update $(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}})$, $\{z_k\}$, $\mathbf{U}_2$ and $\mathbf{F}$ in **Step 2**.
4. – Update $\mathbf{X}$ and $\mathbf{Y}$ in **Step 3**.
5. – Update $\mathbf{S}$ and $\mathbf{W}$ in **Step 4**.
6. – Update the iteration number: $i = i + 1$.
7. **Until** the difference between successive values of the objective function is less than ϵ_1 or the maximum number of iterations is reached, i.e., $i \geq N_{\max}$.

---

update the dual variables as

$$\boldsymbol{\lambda}_1^{m+1} = \boldsymbol{\lambda}_1^m + \frac{1}{\rho^m} \left(\mathbf{U}_2 \mathbf{W} \mathbf{U}_1 - \mathbf{Y} \right),$$

$$\boldsymbol{\lambda}_2^{m+1} = \boldsymbol{\lambda}_2^m + \frac{1}{\rho^m} \left(\mathbf{V} \mathbf{F} - \mathbf{S} \right),$$

$$\boldsymbol{\lambda}_3^{m+1} = \boldsymbol{\lambda}_3^m + \frac{1}{\rho^m} \left(\mathbf{Y} \mathbf{G} \mathbf{S} - \tilde{\mathbf{X}} \right),$$

$$\boldsymbol{\lambda}_4^{m+1} = \boldsymbol{\lambda}_4^m + \frac{1}{\rho^m} \left(\mathbf{H}_R \mathbf{X} - \bar{\mathbf{X}} \right),$$

$$\boldsymbol{\lambda}_5^{m+1} = \boldsymbol{\lambda}_5^m + \frac{1}{\rho^m} \left(\mathbf{X} - \tilde{\mathbf{X}} \right),$$

$$\boldsymbol{\lambda}_6^{m+1} = \boldsymbol{\lambda}_6^m + \frac{1}{\rho^m} \left(\sigma_R \mathbf{H}_R \mathbf{Y} - \bar{\mathbf{Y}} \right), \tag{49}$$

$$\boldsymbol{\lambda}_7^{m+1} = \boldsymbol{\lambda}_7^m + \frac{1}{\rho^m} \left(\sigma_R \mathbf{Y} - \tilde{\mathbf{Y}} \right),$$

$$\boldsymbol{\lambda}_8^{m+1} = \boldsymbol{\lambda}_8^m + \frac{1}{\rho^m} \left(\tilde{\mathbf{S}} - \mathbf{S} \right),$$

$$\boldsymbol{\lambda}_{9,k}^{m+1} = \boldsymbol{\lambda}_{9,k}^m + \frac{1}{\rho^m} \left(\hat{\mathbf{X}}_k - \mathbf{X} \right),$$

$$\boldsymbol{\lambda}_{10,k}^{m+1} = \boldsymbol{\lambda}_{10,k}^m + \frac{1}{\rho^m} \left(\hat{\mathbf{Y}}_k - \mathbf{Y} \right), \ k \in \mathcal{K}.$$

We define a constraint violation indicator h as
$$h = \max\{ \|\mathbf{U}_2 \mathbf{W} \mathbf{U}_1 - \mathbf{Y}\|, \|\mathbf{V}\mathbf{F} - \mathbf{S}\|, \|\mathbf{Y}\mathbf{G}\mathbf{S} - \tilde{\mathbf{X}}\|,$$
$$\|\mathbf{H}_R \mathbf{X} - \bar{\mathbf{X}}\|, \|\mathbf{X} - \tilde{\mathbf{X}}\|, \|\sigma_R \mathbf{H}_R \mathbf{Y} - \bar{\mathbf{Y}}\|,$$
$$\|\sigma_R \mathbf{Y} - \tilde{\mathbf{Y}}\|, \|\tilde{\mathbf{S}} - \mathbf{S}\|, \|\hat{\mathbf{X}}_k - \mathbf{X}\|, \tag{50}$$
$$\|\hat{\mathbf{Y}}_k - \mathbf{Y}\|, \forall k \in \mathcal{K} \}.$$

We terminate the algorithm when $h < \epsilon_2$, where ϵ_2 denotes the tolerance of accuracy of the outer loop. Furthermore, based on the convergence analysis in [26] and the equivalence between problems (13) and (17), we can conclude that the proposed PDD-based joint optimization algorithm converges to the set of stationary solutions of problem (13). To better guide the reader, the internal structure of the PDD-based algorithm is exposed in the form of a flow-chart in Fig. 3. In particular, in the inner loop of the algorithm, the optimized variables are divided into four blocks, which are updated sequentially. Within each block however, the variables can be optimized in parallel.
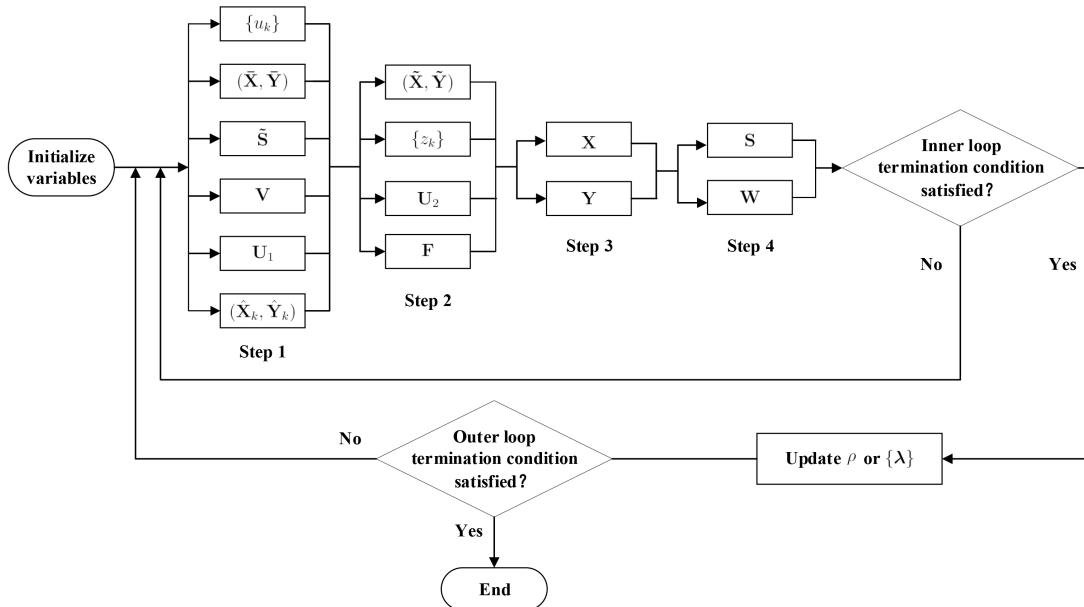
E. Summary of the proposed PDD-based algorithm

In each iteration of the PDD-based algorithm, after implementing the CCCP optimization algorithm in **Algorithm 2**, we adjust the penalty parameter ρ and the dual variables according to Steps 4–10 in **Algorithm 1**. Specifically, we decrease the penalty parameter by updating $\rho \leftarrow c\rho$, where $0 < c < 1$, and

Fig. 3: The flow chart exposing the internal structure of the proposed PDD-based algorithm.

V. PRACTICAL DESIGN CONSIDERATIONS

In this section, we first extend the proposed joint hybrid transceiver design algorithm to the case where finite resolution phase shifters are employed in the FD mmWave MIMO relay system. We then present the detailed computational complexity analysis of the proposed PDD-based algorithms for both cases of infinite and finite resolution phase shifters.

A. Transceiver design with finite resolution phase shifters

In practice, the mmWave hardware for accurate phase control could be very expensive. Thus, we need to consider hybrid beamforming with finite resolution phase shifters, i.e., when the set of candidate phases for each phase shifter is finite.

Let $\mathcal{F}$ denote the finite codebook of phases, with cardinality $|\mathcal{F}| = 2^B$, where B denotes the number of bits used to quantize the phases. The corresponding problem can be formulated as (51). The proposed joint hybrid transceiver design algorithm introduced in Section IV can be straightforwardly adapted to the above case with finite phase resolution. The modification lies strictly in the optimization of $\mathbf{V}$, $\mathbf{U}_1$, and $\mathbf{U}_2$ in the subproblems of **Step 1** and **Step 2**, which in each case amounts to a quadratic optimization with unit modulus constraint. Therefore, by taking the constraint of finite discrete phases into account, (28) can be reformulated as

$$\min_{\mathbf{V}(i,j)\in\mathcal{F}} \text{Tr}(\mathbf{V}^H\mathbf{V}\mathbf{C}_V) - 2\Re\{\text{Tr}(\mathbf{V}^H\mathbf{B}_V)\}, \qquad (52)$$

where we take the optimization of $\mathbf{V}$ as an example. Problem (52) can be tackled by using the one-iteration block coordinate descent type algorithm with one-dimensional exhaustive search over $\mathcal{F}$ [28].

B. Computational complexity

In the following, we analyze the computational complexity of the proposed PDD-based joint hybrid transceiver design algorithm, where the number of complex multiplications is used as a measure of complexity. We first focus our attention on the computational complexity of updating $(\hat{\mathbf{X}}_k, \hat{\mathbf{Y}}_k), \forall k$ and $\mathbf{W}$ in **Step 1** and **Step 4**, respectively, of the algorithm introduced in Subsection IV-D. For the subproblem with respect to $(\hat{\mathbf{X}}_k, \hat{\mathbf{Y}}_k), \forall k$, since $\mathbf{a}_k$ $\hat{\mathbf{h}}_k$, $\boldsymbol{\lambda}_{9,k}$ and $\boldsymbol{\lambda}_{10,k}$ are invariant in the optimization, we can precalculate the corresponding terms such as $(\mathbf{a}_j\mathbf{a}_j^T) \otimes \hat{\mathbf{h}}_k\hat{\mathbf{h}}_k^H$. Notwithstanding the computation of these constant terms, the complexity of this subproblem is dominated by the matrix inversion and the bisection method used to search the Lagrangian parameter v_k. The number of iterations for the latter is $\log_2(\frac{\theta_{0,s}}{\theta_s})$, where $\theta_{0,s}$ is the initial interval size and θ_s is the tolerance. The calculation of $\hat{\mathbf{X}}_k$ and $\hat{\mathbf{Y}}_k$ involves the inverse of a matrix based on Gauss-Jordan elimination, with complexity[4] $\mathcal{O}\left(N_r^3\right)$. Hence, the overall complexity of updating $(\hat{\mathbf{X}}_k, \hat{\mathbf{Y}}_k), \forall k$ is $\mathcal{O}\left(N_r^3 + \log_2(\frac{\theta_{0,s}}{\theta_s})\right)$. For the subproblem with respect to $\mathbf{W}$, the computational complexity is dominated by the matrix operations in (48), and is given by $\mathcal{O}(N_{RF}^6) + \mathcal{O}(N_{RF}N_r^2)$.

By following similar steps as in the above analysis for the subproblems with respect to $(\hat{\mathbf{X}}_k, \hat{\mathbf{Y}}_k), \forall k$ and $\mathbf{W}$, we can obtain the computational complexity for the other subproblems, as summarized in Table I. In addition, note that the proposed one-iteration block coordinate descent type algorithm shown in [28] has a complexity of $\mathcal{O}(n^2)$, where n denotes the number of input variables. Therefore, by retaining dominant terms the

---

[4]Due to the block diagonal structure of matrix $\mathbf{J}_x(v_k)$ (see (66) in Appendix C), the computation of $\hat{\mathbf{X}}_k$ can be decomposed into K parallel components, which requires $\mathcal{O}\left(N_r^3\right)$ complexity.

$$\max_{\mathbf{F},\mathbf{W},\mathbf{V},\mathbf{U}_1,\mathbf{U}_2} \quad \sum_{k=1}^{K} \log \left(1 + \frac{|\hat{\mathbf{h}}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1 \mathbf{G} \mathbf{V} \mathbf{f}_k|^2}{\sum_{k'\neq k}^{K} |\hat{\mathbf{h}}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1 \mathbf{G} \mathbf{V} \mathbf{f}_{k'}|^2 + \sigma_R^2 \|\hat{\mathbf{h}}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1\|^2 + N_k} \right) \tag{51a}$$

$$\text{s.t.} \quad (13b) - (13e) \tag{51b}$$

$$\mathbf{V}(i,j) \in \mathcal{F}, \ \mathbf{U}_1(i,j) \in \mathcal{F}, \ \mathbf{U}_2(i,j) \in \mathcal{F}, \ \forall i,j. \tag{51c}$$

overall complexity of the proposed joint hybrid transceiver design is

$$\mathcal{O}\left(I_1 I_2 \left(N_r^3 + \log_2(\frac{\theta_{0,s}}{\theta_s}) + (N_{RF}N_r)^2 + N_{RF}^6 \right) \right), \tag{53}$$

where I_1 and I_2 denote the maximum number of iterations for the outer and inner loops, respectively. Similarly, we can find that the computational complexity of the proposed design algorithm with finite resolution phase shifters is

$$\mathcal{O}\left(I_1 I_2 \left(N_r^3 + \log_2(\frac{\theta_{0,s}}{\theta_s}) + (N_{RF}N_r + 2^B)N_{RF}N_r + N_{RF}^6 \right) \right). \tag{54}$$

VI. SIMULATION RESULTS

In this section, we evaluate the performance of the proposed PDD-based joint hybrid transceiver design algorithm for mmWave FD MIMO relay systems by means of computer simulations.

In the simulations, we employ a widely used mmWave channel model with uniform linear antenna array configuration, as in e.g., [43]–[46]. Specifically, based on the Saleh-Valenzuela model [47], [48], the channel matrix between the BS and RS[5] can be expressed as

$$\mathbf{G} = \sqrt{M_t M_r} L_p \sum_{l=1}^{L_p} \alpha_l \mathbf{a}_r(\theta_l^r) \mathbf{a}_t(\theta_l^t) \tag{55}$$

where L_p is the number of distinguishable paths, $\alpha_l \sim \mathcal{CN}(0,1)$ is the complex gain of the l-th path, $\mathbf{a}_r(\theta_l^r)$ and $\mathbf{a}_t(\theta_l^t)$ are the receive and transmit array response vectors, where $\theta_l^r \in [0, 2\pi)$ and $\theta_l^t \in [0, 2\pi)$ are the azimuth angles of arrival and departure (AoAs and AoDs), respectively. The generic expression for the response vector is given by

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{M}}[1, e^{jk_o d_a \pi \sin(\theta)}, \ldots, e^{jk_o d_a (M-1)\pi \sin(\theta)}]^T \tag{56}$$

where $k_o = 2\pi/\lambda_o$, λ_o is the wavelength at the operating frequency and d_a is the antenna spacing. The SI channel matrix $\mathbf{H}_R$ is generated based on the model described in [49].

We set the desired per user transmission rate in (13e) to $\gamma_k = 0.6$ bits per channel use. We assume that the BS and the RS are equipped with the same number of RF chains, i.e., $R_1 = R_2 \triangleq N_{RF}$. The input SNR at the relay is defined as $\text{SNR}_1 \triangleq \frac{P_2}{N_r \sigma^2}$ while the SNR at the destination is defined as $\text{SNR}_2 \triangleq \frac{P_1}{K \sigma_k^2}$. In the simulations, we let $\text{SNR}_1 = 15$ dB and vary SNR_2 for simplicity. All the results are obtained

[5]Here we provide the detailed mmWave channel matrix $\mathbf{G}$ of the link between the BS and the RS, while the channel matrix $\mathbf{H}$ between the BS and the users can be obtained similarly.
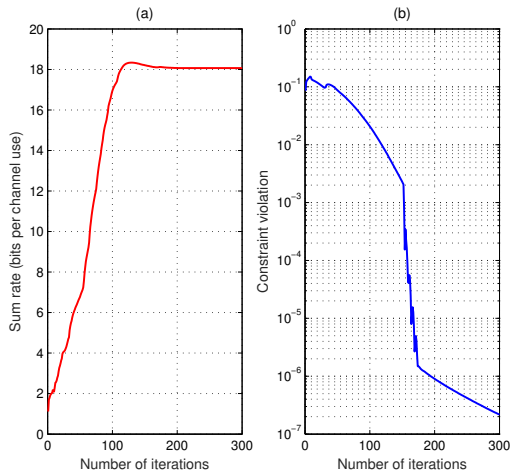

Fig. 4: (a) Sum rate and (b) constraint violation versus the number of iterations for the proposed PDD-based joint hybrid transceiver design algorithm.

by averaging over 1000 independent Monte Carlo runs. The tolerance parameters for the proposed algorithm are chosen as $\epsilon_1 = \epsilon_2 = 10^{-4}$ while $N_{max} = 400$. The parameter c in the PDD outer loop (see **Algorithm 1**) is set to $c = 0.95$. We consider different variations of the proposed design algorithm as well as other benchmark approaches, as indicated by the following terminology:

- PDD: The proposed PDD-based joint hybrid transceiver design algorithm.
- B bits: The proposed algorithm with finite resolution phase shifters employing B bits, i.e., with a finite codebook of size[6] 2^B.
- $B = \infty$: The transceiver design with infinite resolution phase shifters, i.e. without considering phase quantization.
- FD: Full-duplex mode of operation.
- HD: Half-duplex mode of operation.
- Robust: The proposed PDD-based design relying on worst-case sum rate maximization.
- Nonrobust: The PDD-based design without taking the CSI errors into account.

Let us first study the convergence of the proposed PDD-based robust joint hybrid transceiver algorithm versus the number of outer iterations. Here, we set $N_t = 8$, $N_r = 16$, $N_{RF} = 4$, $K = 3$, $\sigma_e^2 = 0.04$ and $\text{SNR}_2 = 10$ dB. In Fig. 4 (a), we show the sum rate performance versus the number of iterations for the proposed algorithm, which is seen to

[6]The available phase values are uniformly distributed around the unit circle.

TABLE I: Complexity analysis of the proposed CCCP algorithm in the inner loop

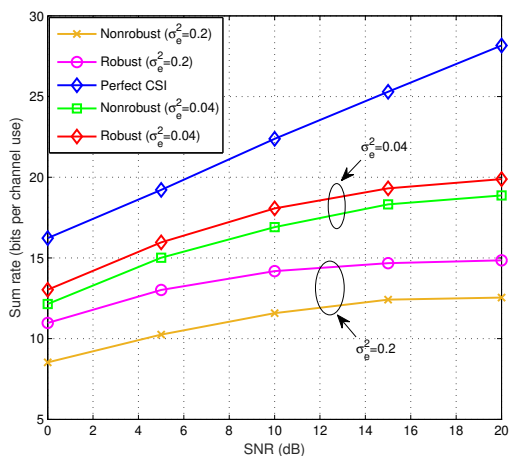| Subproblem		Complexity
Step 1	$\{u_k\}$	$\mathcal{O}(N_r^2)$
	$(\mathbf{X}, \mathbf{Y})$	$\mathcal{O}(N_r^3)$
	$\bar{\mathbf{S}}$	$\mathcal{O}(N_t K)$
	$\mathbf{V}$	$\mathcal{O}((N_{RF} N_t)^2)$
	$\mathbf{U}_1$	$\mathcal{O}((N_{RF} N_r)^2)$
	$\{\hat{\mathbf{X}}_k, \hat{\mathbf{Y}}_k\}$	$\mathcal{O}\left(N_r^3 + \log_2(\frac{\theta_{0,s}}{\theta_s})\right)$
Step 2	$(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}})$	$\mathcal{O}(N_r^2 N_t)$
	$\{z_k\}$	$\mathcal{O}(N_r^2)$
	$\mathbf{U}_2$	$\mathcal{O}((N_{RF} N_r)^2)$
	$\mathbf{F}$	$\mathcal{O}(N_{RF}^2 N_t)$
Step 3	$\mathbf{X}$	$\mathcal{O}(N_r^3)$
	$\mathbf{Y}$	$\mathcal{O}(N_r^3)$
Step 4	$\mathbf{S}$	$\mathcal{O}(N_t^3)$
	$\mathbf{W}$	$\mathcal{O}(N_{RF}^6) + \mathcal{O}(N_{RF} N_r^2)$

Fig. 5: Comparison of sum rate performance versus SNR for the proposed robust and nonrobust joint hybrid transceiver design algorithms.

Fig. 6: Sum rate performance versus the number of RF chains for the proposed PDD-based FD and conventional HD design algorithms.

convergence within less than 200 iterations.[7] Fig. 4 (b) shows the corresponding value of the constraint violation indicator, i.e., penalty term (50), versus the iteration number. We find that the penalty terms decrease to a value below 10^{-6} after 200 iterations, which supports our claim that the proposed PDD-based algorithm can effectively tackle the equality constraints.

Next, we investigate the robustness of the proposed PDD-based joint hybrid transceiver design against CSI errors. The PDD-based joint design without taking CSI errors into account is also implemented for comparison. Fig. 5 shows the sum rate performance of these schemes versus SNR_2, where we set $N_t = 8$, $N_r = 16$, $N_{RF} = 4$ and $K = 3$. From the results, we can see that the proposed robust joint transceiver design outperforms its nonrobust counterpart by a significant margin, where the performance degradation of the nonrobust design is due to the CSI mismatch. These results demonstrate the effectiveness of the proposed design approach based on

the worst-case sum rate maximization, as a means to handle CSI errors. In particular, the performance gain with the robust design increases with the variance of the CSI errors. The performance of the perfect CSI case is provided as an upper bound reference.

Fig. 6 shows the sum rate performance versus the number of RF chains for the proposed robust joint transceiver design for FD mmWave relaying versus the conventional transceiver design for HD mmWave relaying, as per [7], [24], [25]. For these experiments we set $\sigma_e^2 = 0.04$, $N_t = N_r = 16$, $K = 3$ and $\text{SNR}_2 = 5$ dB. In Fig. 6, we can see that the sum rate performance of the proposed FD and conventional HD designs increases with the number of RF chains increases. However, compared to the proposed FD scheme, the conventional scheme suffers from a significant performance loss due to the HD operation. The performance of the HD scheme does not increase as dramatically as that of the proposed FD algorithm, we here consider the scenario with a small number of users and the performance gain of the proposed FD algorithm compared to the conventional HD algorithm is almost doubled with the increasing of the number of RF chains. The performance of the fully digital FD scheme is

---

[7]Due to the parameter adjustment in the outer loop of the proposed PDD-based algorithm, the value of the objective function at each iteration does not change monotonically. However, the proposed algorithm does converge to a stationary solution of the original problem, and the equality constraints are guaranteed.
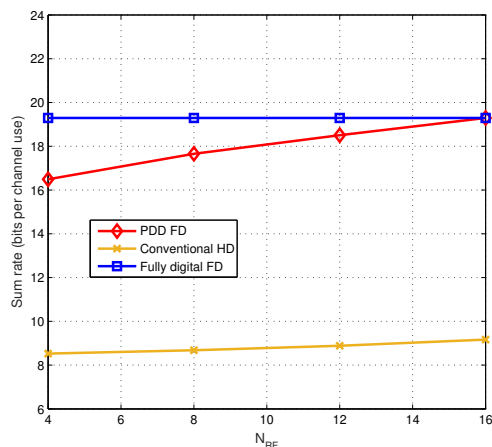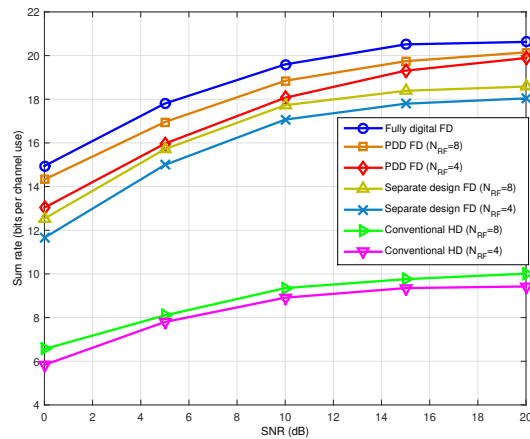
Fig. 7: Sum rate performance versus SNR_2 for the proposed FD and conventional algorithms.

Fig. 8: Sum rate performance versus SNR_2 for the proposed FD and conventional algorithms in the case of large-scale antenna arrays.

provided as a reference. In Fig. 7, we show the sum rate performance as a function of SNR_2 for the proposed FD joint hybrid transceiver design, the FD separate transceiver design[8] and the conventional HD scheme, where we set $N_t = 8$, $N_r = 16$ and $K = 3$. The performance of the fully digital transceiver is provided as a reference. It is observed that the fully digital scheme achieves the best performance, followed by the proposed FD joint transceiver design, the FD separate transceiver design, and the conventional HD scheme. In particular, thanks to the FD operation of the proposed algorithm, the sum rate performance is approximately doubled compared to the HD scheme. Since we jointly optimize the analog and digital beamforming matrices and the proposed algorithm can converge to a stationary solution of the original problem, the performance of the FD algorithm is much better than that of the heuristic separate design. Fig. 8 shows the sum rate performance versus SNR_2 for the case of large-scale antenna arrays, where we set $N_t = N_r = 64$, $N_{RF} = 8$ and $K = 3$. From the results, we note that the proposed design achieves the best performance amongst the systems using a reduced number of RF chains, and this by a significant margin.

In the following, we focus on the performance of the proposed robust FD joint hybrid transceiver design with finite resolution phase shifters, where the PDD-based algorithm introduced in Section IV is modified as explained in Subsection V-A. Fig. 9 shows the convergence behavior of this algorithm using $B = 6$ bits for phase quantization, where we set $N_t = 8$, $N_r = 16$, $N_{RF} = 4$, $K = 3$, $\sigma_e^2 = 0.04$ and $SNR_2 = 10$dB. In particular, Fig. 9 (a) and Fig. 9 (b) illustrate the sum rate and constraint violation versus the number of iterations, respectively. In this case, the proposed algorithm converges within 300 iterations, while the penalty term decreases to a value below 10^{-4}. Fig. 10 depicts the sum rate performance versus SNR_2 for the proposed algorithm with finite resolution phase shifters for different numbers of quantization bits, i.e.,

Fig. 9: (a) Sum rate and (b) constraint violation versus the number of iterations for the proposed joint hybrid transceiver design with finite resolution phase shifters ($B = 6$ bits).

Fig. 10: Sum rate performance versus SNR_2 for the proposed algorithm with finite resolution phase shifters.

---

[8]For the separate design, the analog beamforming matrices are determined by using the channel matching method and the baseband beamforming matrices are optimized based on the proposed PDD-based algorithm.
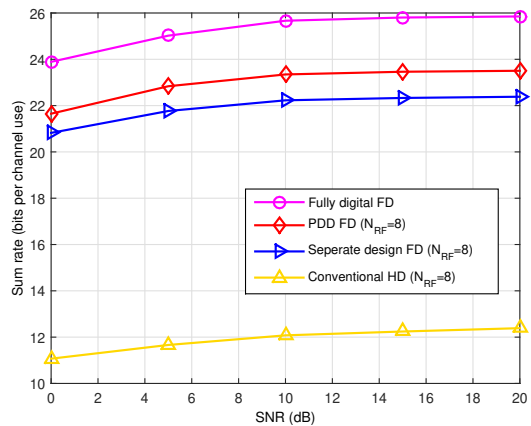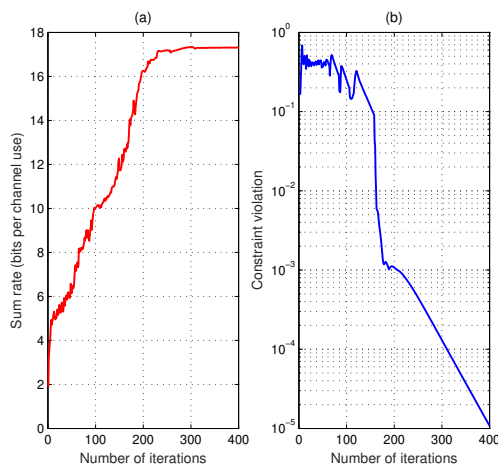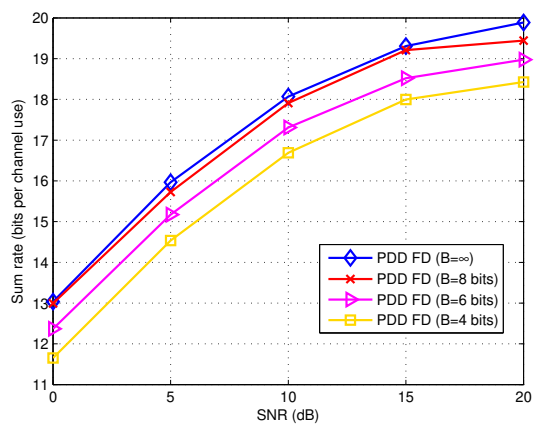
$B = 4$, 6, and 8. In this simulation, we set $N_t = 8$, $N_r = 16$, $N_{RF} = 4$, $K = 3$ and $\sigma_e^2 = 0.04$, while the performance of the proposed FD joint hybrid transceiver design with infinite resolution phase shifters is provided as a reference. We first note that due to the finite resolution of phase shifters, the sum rate performance degrades monotonically when the number B of quantization bits is reduced, allowing a trade-off between complexity and transmit power. For instance, for a sum rate level of 18 bits per channel use, the proposed algorithm with $B = 8$ bits can save up to 4 dB in transmit power when compared to that with $B = 4$ bits. We also observe that the performance of the proposed PDD-based algorithm with finite resolution phase shifters nearly achieves that of the infinite resolution phase shifters for $B = 8$ quantization bits, which demonstrates the effectiveness of the proposed design algorithm in practice.

VII. CONCLUSIONS

In this paper, we have investigated the robust joint hybrid beamforming design for a multiuser mmWave FD relay system. To tackle this challenging problem, where the design variables are highly coupled through the objective and constraint functions, a novel PDD-based algorithm has been proposed to jointly optimize the BS digital and analog beamforming matrices, the RS receive and transmit analog beamforming matrices, and the RS digital AF beamforming matrix under the transmit power constraints, the residual self-interference power constraint at the RS, the per user QoS constraint, along with the unit modulus constraints on the elements of analog beamforming matrices. The proposed PDD-based algorithm performs double-loop iterations: the inner loop update the optimization variables in a block coordinate descent fashion; while the outer loop adjusts the Lagrange multipliers and penalty parameter. The convergence of the proposed PDD-based algorithm has been discussed, along with its extension to the case of finite resolution phase shifters. A detailed computational complexity analysis was also provided. Our simulation results have demonstrated that the proposed PDD-based algorithm for robust joint design of hybrid FD relay systems can achieve better performance than the conventional hybrid beamforming algorithms.

APPENDIX A
DERIVATION OF (12)

The Woodbury matrix identity is given by [50]

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{DA}^{-1}, \tag{57}$$

where $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ denote matrices of appropriate dimensions. Let us assign $\mathbf{A} \leftarrow \mathbf{I}$, $\mathbf{B} \leftarrow \hat{\mathbf{h}}_k$, $\mathbf{C} \leftarrow \Upsilon_k^{-1}$ and $\mathbf{D} \leftarrow \mathbf{h}_k^H \mathbf{Q}_k$, then by using (57), we have

$$\mathbf{I} - \frac{\hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H \mathbf{Q}_k}{\hat{\mathbf{h}}_k^H \mathbf{Q}_k \hat{\mathbf{h}}_k + \Upsilon_k} = \left(\mathbf{I} + \frac{\hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H \mathbf{Q}_k}{\Upsilon_k} \right)^{-1}. \tag{58}$$

Therefore, the left hand side (LHS) of (10) can be rewritten as

$$\frac{1}{2} \log \det \left(2\pi e \mathbf{Q}_k \left(\mathbf{I} - \frac{\hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H \mathbf{Q}_k}{\hat{\mathbf{h}}_k^H \mathbf{Q}_k \hat{\mathbf{h}}_k + \Upsilon_k} \right) \right)$$
$$= \frac{1}{2} \log \det \left(2\pi e \mathbf{Q}_k \left(\mathbf{I} + \frac{\hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H \mathbf{Q}_k}{\Upsilon_k} \right)^{-1} \right).$$
$$= \frac{1}{2} \log \det \left(2\pi e \mathbf{Q}_k \right) - \frac{1}{2} \log \det \left(\mathbf{I} + \frac{\hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H \mathbf{Q}_k}{\Upsilon_k} \right). \tag{59}$$

Using (59) and invoking the identity $\det(\mathbf{I} + \mathbf{AB}) = \det(\mathbf{I} + \mathbf{BA})$, (10) can be rewritten as

$$\mathcal{H}(\mathbf{p}_k|y_k,\hat{\mathbf{H}}) \leq \frac{1}{2} \log \det \left(2\pi e \mathbf{Q}_k \left(\mathbf{I} - \frac{\hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H \mathbf{Q}_k}{\hat{\mathbf{h}}_k^H \mathbf{Q}_k \hat{\mathbf{h}}_k + \Upsilon_k} \right) \right)$$
$$= \frac{1}{2} \log \det \left(2\pi e \mathbf{Q}_k \right)$$
$$- \frac{1}{2} \log \det \left(1 + \frac{\hat{\mathbf{h}}_k^H \mathbf{Q}_k \hat{\mathbf{h}}_k}{\Upsilon_k} \right). \tag{60}$$

Finally, substituting (60) into (8) yields

$$\mathcal{H}(\mathbf{p}_k; y_k|\hat{\mathbf{H}}) \geq \frac{1}{2} \log \left(1 + \frac{|\hat{\mathbf{h}}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1 \mathbf{G} \mathbf{V} \mathbf{f}_k|^2}{\Upsilon_k} \right). \tag{61}$$

APPENDIX B
PROOF OF EQUIVALENCE BETWEEN (13) AND (17)

Note that u_k and z_k only appear in the objective function of (17). Hence, by fixing the other variables, the optimum u_k for minimizing (17) is given in (62). It can be seen that the objective function of (17) is convex with respect to z_k when fixing the other variables. Hence, the optimum z_k in (17) can be obtained based on the first order optimality condition as follows,

$$z_k = e_k^{-1}, \forall k. \tag{63}$$

By substituting (62) and (63) in (17), we have the equivalent optimization problem shown in (64). It is readily seen that (64c) is equivalent to (13e). Finally, we obtain problem (13). This completes the proof.

APPENDIX C
SOLUTION TO PROBLEM (32)

Note that (32) can be solved in closed form based on the Lagrange multipliers method. By attaching a Lagrange multiplier v_k to constraint (22), we then obtain the following Lagrange function

$$L(\hat{\mathbf{X}}_k, \hat{\mathbf{Y}}_k, v_k) \triangleq \sum_{k=1}^{K} (\|\hat{\mathbf{X}}_k - \mathbf{X} + \rho\boldsymbol{\lambda}_{9,k}\|^2 + \|\hat{\mathbf{Y}}_k - \mathbf{Y} + \rho\boldsymbol{\lambda}_{10,k}\|^2))$$
$$+ v_k (\sum_{j \neq k}^{K} |\hat{\mathbf{h}}_k^H \hat{\mathbf{X}}_k \mathbf{a}_j|^2 + \sigma_R^2 \|\hat{\mathbf{h}}_k^H \hat{\mathbf{Y}}_k\|^2$$
$$+ \sigma_e^2 \big(\sum_{j=1}^{K} \|\hat{\mathbf{X}}_k \mathbf{a}_j\|^2 + \sigma_R^2 \|\hat{\mathbf{Y}}_k\|^2 \big) + \sigma_k^2$$
$$+ \frac{1}{\gamma_k'} |\hat{\mathbf{h}}_k^H \hat{\mathbf{X}}_k^{(i)} \mathbf{a}_k|^2 - \frac{2}{\gamma_k'} \Re\{\mathbf{a}_k^H \hat{\mathbf{X}}_k^{(i)H} \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H \hat{\mathbf{X}}_k \mathbf{a}_k\}) \tag{65}$$

$$u_k = \frac{\mathbf{f}_k^H \mathbf{V}^H \mathbf{G}^H \mathbf{U}_1^H \mathbf{W}^H \mathbf{U}_2^H \hat{\mathbf{h}}_k}{\sum_{j=1}^K |\hat{\mathbf{h}}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1 \mathbf{G} \mathbf{f}_j|^2 + \sigma_R^2 \|\hat{\mathbf{h}}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1\|^2 + N_k}, \quad \forall k. \tag{62}$$

$$\max_{\mathbf{F},\mathbf{W},\mathbf{U}_1,\mathbf{U}_2,\mathbf{V}} \quad \sum_{k=1}^K \log\left(1 + \frac{|\hat{\mathbf{h}}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1 \mathbf{G} \mathbf{f}_k|^2}{\sum_{k' \neq k}^K |\hat{\mathbf{h}}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1 \mathbf{G} \mathbf{f}_{k'}|^2 + \sigma_R^2 \|\hat{\mathbf{h}}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1\|^2 + N_k}\right) \tag{64a}$$

$$\text{s.t.} \quad (13b) - (13d),\ (13f) \tag{64b}$$

$$\frac{|\hat{\mathbf{h}}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1 \mathbf{G} \mathbf{f}_k|^2}{\sum_{j \neq 1}^K |\hat{\mathbf{h}}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1 \mathbf{G} \mathbf{f}_j|^2 + \sigma_R^2 \|\hat{\mathbf{h}}_k^H \mathbf{U}_2 \mathbf{W} \mathbf{U}_1\|^2 + N_k} \geq \gamma_k'. \tag{64c}$$

By vectorizing $\hat{\mathbf{X}}_k$ and examining the first order optimality condition of $L(\hat{\mathbf{X}}_k, \hat{\mathbf{Y}}_k, v_k)$ with respect to $\mathrm{vec}(\hat{\mathbf{X}}_k(v_k))$, we obtain

$$\mathrm{vec}(\hat{\mathbf{X}}_k(v_k)) = \mathbf{J}_x(v_k)^{-1} \mathbf{r}_x(v_k), \tag{66}$$

where $\mathbf{J}_x(v_k) \triangleq \mathbf{I} + v_k\left(\sum_{j \neq k}^K (\mathbf{a}_j \mathbf{a}_j^T) \otimes (\hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H) + \sigma_e^2 \mathbf{I}\right)$ and $\mathbf{r}_x(v_k) \triangleq \mathrm{vec}(\mathbf{X} - \rho \boldsymbol{\lambda}_{9,k} + \frac{v_k}{\gamma_k'} \mathrm{vec}(\hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H \hat{\mathbf{X}}_k^{(i)} \mathbf{a}_k \mathbf{a}_k^T))$.

By examining the first order optimality condition of $L(\hat{\mathbf{X}}_k, \hat{\mathbf{Y}}_k, v_k)$ with respect to $\hat{\mathbf{Y}}_k$, we have

$$\hat{\mathbf{Y}}_k(v_k) = \mathbf{J}_y(v_k)^{-1} \mathbf{R}_y, \tag{67}$$

where $\mathbf{J}_y(v_k) \triangleq \mathbf{I} + v_k \sigma_R^2 (\hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H + \sigma_e^2 \mathbf{I})$ and $\mathbf{R}_y \triangleq \mathbf{Y} - \rho \boldsymbol{\lambda}_{10,k}$.

The optimal multiplier v_k should be determined such that the complementarity slackness condition of the constraint is satisfied.

Let us define

$$P_k(\hat{\mathbf{X}}_k, \hat{\mathbf{Y}}_k) \triangleq \sum_{j \neq k}^K |\hat{\mathbf{h}}_k^H \hat{\mathbf{X}}_k \mathbf{a}_j|^2 + \sigma_R^2 \|\hat{\mathbf{h}}_k^H \hat{\mathbf{Y}}_k\|^2$$
$$+ \sigma_e^2 \left(\sum_{j=1}^K \|\hat{\mathbf{X}}_k \mathbf{a}_j\|^2 + \sigma_R^2 \|\hat{\mathbf{Y}}_k\|^2\right) + \sigma_k^2$$
$$+ \frac{1}{\gamma_k'} |\hat{\mathbf{h}}_k^H \hat{\mathbf{X}}_k^{(i)} \mathbf{a}_k|^2 - \frac{2}{\gamma_k'} \Re\{\mathbf{a}_k^H \hat{\mathbf{X}}_k^{(i)H} \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H \hat{\mathbf{X}}_k \mathbf{a}_k\} \tag{68}$$

When $P_k(\hat{\mathbf{X}}_k(0), \hat{\mathbf{Y}}_k(0)) \leq 0$, we have the optimal $\hat{\mathbf{X}}_k = \hat{\mathbf{X}}_k(0)$ and $\hat{\mathbf{Y}}_k = \hat{\mathbf{Y}}_k(0)$, otherwise we must have $P_k(\hat{\mathbf{X}}_k(v_k), \hat{\mathbf{Y}}_k(v_k)) = 0$, which results in the following equation:

$$\mathbf{r}_x(v_k)^H \mathbf{J}_x(v_k)^{-1} \tilde{\mathbf{U}} \mathbf{J}_x(v_k)^{-1} \mathbf{r}_x(v_k)$$
$$+ \mathrm{Tr}(\mathbf{R}_y^H \mathbf{J}_y(v_k)^{-1} \hat{\mathbf{U}} \mathbf{J}_y(v_k)^{-1} \mathbf{R}_y)$$
$$- \frac{1}{\gamma_k'}\left(\mathbf{q}_k^H \mathbf{J}_x(v_k)^{-1} \mathbf{r}_x(v_k) + \mathbf{r}_x(v_k)^H \mathbf{J}_x(v_k)^{-1} \mathbf{q}_k\right)$$
$$+ t_k = 0, \tag{69}$$

where $\tilde{\mathbf{U}} \triangleq \sum_{j \neq k}^K (\mathbf{a}_j \mathbf{a}_j^T) \otimes (\hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H) + \sigma_e^2 \mathbf{I}$, $\hat{\mathbf{U}} \triangleq \sigma_R^2 (\hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H + \sigma_e^2 \mathbf{I})$, $\mathbf{q}_k \triangleq \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H \hat{\mathbf{X}}_k^{(i)} \mathbf{a}_k \mathbf{a}_k^T$ and $t_k \triangleq \sigma_k^2 + \frac{1}{\gamma_k'} |\hat{\mathbf{h}}_k^H \hat{\mathbf{X}}_k^{(i)} \mathbf{a}_k|^2$.

Furthermore, (69) is equivalent to

$$\mathrm{Tr}(\mathbf{J}_x(v_k)^{-1} \tilde{\mathbf{U}} \mathbf{J}_x(v_k)^{-1} \mathbf{r}_x(v_k) \mathbf{r}_x(v_k)^H)$$
$$+ \mathrm{Tr}(\mathbf{J}_y(v_k)^{-1} \hat{\mathbf{U}} \mathbf{J}_y(v_k)^{-1} \mathbf{R}_y \mathbf{R}_y^H)$$
$$- \mathrm{Tr}(\frac{1}{\gamma_k'}(\mathbf{J}_x(v_k)^{-1} \mathbf{r}_x(v_k) \mathbf{q}_k^H$$
$$+ \mathbf{J}_x(v_k)^{-1} \mathbf{q}_k \mathbf{r}_x(v_k)^H)) + t_k = 0. \tag{70}$$

Due to the fact that

$$\mathbf{J}_x(v_k) = \mathbf{I} + v_k \tilde{\mathbf{U}} = \tilde{\mathbf{U}}^{\frac{1}{2}}(\tilde{\mathbf{U}}^{-1} + v_k \mathbf{I})\tilde{\mathbf{U}}^{\frac{1}{2}}, \tag{71}$$

$$\mathbf{J}_y(v_k) = \mathbf{I} + v_k \hat{\mathbf{U}} = \hat{\mathbf{U}}^{\frac{1}{2}}(\hat{\mathbf{U}}^{-1} + v_k \mathbf{I})\hat{\mathbf{U}}^{\frac{1}{2}}, \tag{72}$$

(70) can be rewritten as

$$\mathrm{Tr}((\tilde{\mathbf{U}}^{-1} + v_k \mathbf{I})^{-2} \tilde{\mathbf{U}}^{-\frac{1}{2}} \hat{\mathbf{r}}(v_k) \hat{\mathbf{r}}(v_k)^H \tilde{\mathbf{U}}^{-\frac{1}{2}})$$
$$+ \mathrm{Tr}((\hat{\mathbf{U}}^{-1} + v_k \mathbf{I})^{-2} \hat{\mathbf{U}}^{-\frac{1}{2}} \mathbf{R}_y \mathbf{R}_y^H \hat{\mathbf{U}}^{-\frac{1}{2}})$$
$$- \mathrm{Tr}(\frac{1}{\gamma_k'}((\tilde{\mathbf{U}}^{-1} + v_k \mathbf{I})^{-1} \tilde{\mathbf{U}}^{-\frac{1}{2}}(\mathbf{r}_x(v_k) \mathbf{q}_k^H$$
$$+ \mathbf{q}_k \mathbf{r}_x(v_k)^H) \tilde{\mathbf{U}}^{-\frac{1}{2}})) + t_k = 0. \tag{73}$$

Finally, (73) can be equivalently expressed as

$$\sum_{i=1}^{KN_r} \frac{d_i(v_k)}{(s_i + v_k)^2} + \sum_{i=1}^{N_r} \frac{\tilde{d}_i}{(\tilde{s}_i + v_k)^2} - \frac{1}{\gamma_k'} \sum_{i=1}^{N_r^2} \frac{g_i(v_k)}{s_i + v_k} + t_k = 0, \tag{74}$$

where

$$d_i(v_k) = [\tilde{\mathbf{V}}^H \tilde{\mathbf{U}}^{-\frac{1}{2}} \mathbf{r}_x(v_k) \mathbf{r}_x(v_k)^H \tilde{\mathbf{U}}^{-\frac{1}{2}} \tilde{\mathbf{V}}]_{i,i},$$
$$\tilde{d}_i = [\hat{\mathbf{V}}^H \hat{\mathbf{U}}^{-\frac{1}{2}} \mathbf{R}_y \mathbf{R}_y^H \hat{\mathbf{U}}^{-\frac{1}{2}} \hat{\mathbf{V}}]_{i,i}, \tag{75}$$
$$g_i(v_k) = [\tilde{\mathbf{V}}^H \tilde{\mathbf{U}}^{-\frac{1}{2}} (\mathbf{r}(v_k) \mathbf{q}_k^H + \mathbf{q}_k \mathbf{r}(v_k)^H) \tilde{\mathbf{U}}^{-\frac{1}{2}} \tilde{\mathbf{V}}]_{i,i}.$$

Here, $\tilde{\mathbf{V}}$ denotes a unitary matrix consisting of the eigenvectors of $\tilde{\mathbf{U}}^{-1}$, $\hat{\mathbf{V}}$ denotes a unitary matrix consisting of the eigenvectors of $\hat{\mathbf{U}}^{-1}$, and s_i denotes the corresponding eigenvalues of $\tilde{\mathbf{U}}^{-1}$, so that $\tilde{\mathbf{U}}^{-1} = \tilde{\mathbf{V}}^H \mathrm{diag}\{s_1, s_2, \ldots, s_{KN_r}\}\tilde{\mathbf{V}}$, and $\tilde{s}_i$ denotes the corresponding eigenvalues of $\hat{\mathbf{U}}^{-1}$, so that $\hat{\mathbf{U}}^{-1} = \hat{\mathbf{V}}^H \mathrm{diag}\{\tilde{s}_1, \tilde{s}_2, \ldots, \tilde{s}_{N_r}\}\hat{\mathbf{V}}$. Note that the optimal v_k must be positive, hence (74) can be solved using a one dimensional search. Finally, by substituting the optimal v_k in (66) and (67), we obtain the solution for $\{\hat{\mathbf{X}}_k, \hat{\mathbf{Y}}_k\}$.

REFERENCES

[1] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.,* vol. 49, no. 6, pp. 101–107, Jun. 2011.

[2] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimo, and F. Gutierrez,"Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access,* vol. 1, pp. 335–349, May 2013.

[3] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE,* vol. 102, no. 3, pp. 366–385, Mar. 2014.

[4] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.,* vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[5] I. Hemadeh, K. Satyanarayana, M. El-Hajjar, L. Hanzo: Millimeter-Wave Communications: Physical Channel Models, Design Considerations, Antenna Constructions and Link-Budget, *IEEE Communications Surveys & Tutorials,* 2018, Early Access

[6] O. El Ayach, R. W. Heath, Jr., S. Abu-Surra, S. Rajagopal, and Z. Pi, "Low complexity precoding for large millimeter wave MIMO systems," in *Proc. ICC,* Jun. 2012, pp. 3724–3729.

[7] J. Zhang, M. Haardt, I. Soloveychik, and A. Wiesel, "A channel matching based hybrid analog-digital strategy for massive multi-user MIMO downlink systems," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop,* Jul. 2016, pp. 1–5.

[8] X. Zhai, Y. Cai, Q. Shi, M. Zhao, G. Y. Li, and B. Champagne, "Joint transceiver design with antenna selection for large-scale MU-MIMO millimeter-wave systems," *IEEE J. Sel. Areas Commun.,* vol. 35, no. 9, pp. 2085–2096, Sep. 2017.

[9] S. He, J. Wang, Y. Huang, B. Ottersten, and W. Hong, "Codebook-Based Hybrid Precoding for Millimeter Wave Multiuser Systems," *IEEE Trans. Signal Process.,* vol. 65, no. 20, pp. 5289–5304, Oct. 15, 2017.

[10] C. Lin, G. Y. Li, and L. Wang, "Subarray-based coordinated beamforming training for mmWave and Sub-THz Communications," *IEEE J. Sel. Areas Commun.,* vol. 35, no. 9, pp. 2115–2126, Jun. 2017.

[11] Z. Xiao, T. He, P. Xia, and X.-G. Xia "Hierarchical codebook design for beamforming training in millimeter-wave communication," *IEEE Trans. Wireless Commun.,* vol. 15, no. 5, pp. 3380–3392, May 2016.

[12] Z. Xiao, P. Xia, and X.-G. Xia "Codebook design for millimeter-wave channel estimation with hybrid precoding structure," *IEEE Trans. Wireless Commun.,* vol. 16, no. 1, pp. 141–153, Jan. 2017.

[13] X. Liu, Z. Xiao, L. Bai, J. Choi, P. Xia, and X.-G. Xia, "Beamforming based full-duplex for millimeter-wave communication," *Sensors,*, 2016, 16, 1130; doi:10.3390/s16071130

[14] C. Lin and G. Y. Li, "Terahertz communications: an array-of-subarrays solution," *IEEE Commun. Mag.,* vol. 54, no. 12, pp. 124–131, Dec. 2016.

[15] Z. Zhang, K. Long, A. V. Vasilakos, and L. Hanzo, "Full-duplex wireless communications: Challenges, solutions, and future research directions," *Proceedings of the IEEE*, vol. 104, no. 7, pp. 1369–1409, July 2016.

[16] T. Riihonen, S. Werner, and R. Wichman, "Mitigation of loopback self-interference in full-duplex MIMO relays," *IEEE Trans. Signal Process.,* vol. 59, no. 12, pp. 5983-5993, Dec. 2011.

[17] P. Lioliou, M. Viberg, M. Coldrey, and F. Athley, "Self-interference suppression in full-duplex MIMO relays," in *Conf. Rec. 44th Asilomar Signals, Syst. Comput. Conf.,* Pacific Grove, CA, USA, Nov. 2010, pp. 658-662.

[18] D. Bharadia, E.McMilin, and S.Katti, "Full duplex radio", *Proc. ACM Special Interest Group on Data Commun*, Hong Kong, China, Aug. 2013, pp. 375-386.

[19] D. Bharadia and S. Katti, "Full duplex MIMO radios," in *Proc. 11th USENIX Symp. Netw. Syst. Des. Implementation,* Seattle, WA, USA, Apr. 2014, pp. 359-372.

[20] E. Everett, A. Sahai and A. Sabharwal, "Passive self-interference suppression for full-duplex infrastructure nodes," *IEEE Trans. Wireless Commun.,* vol. 13, no. 2, pp. 680–694, Feb. 2014.

[21] H. Suraweera, I. Krikidis, G. Zheng, C. Yuen, and P. Smith, "Low-complexity end-to-end performance optimization in MIMO full-duplex relay", *IEEE Trans. Wireless Commun,* vol. 13, pp. 913-927, Feb. 2014.

[22] Q. Shi, M. Hong, X. Gao, E. Song, Y. Cai, and W. Xu, "Joint source-relay design for full-duplex MIMO AF relay systems", *IEEE Trans. Signal Process.,* vol. 64, no. 23, pp. 6118-6131, Dec. 1, 2016.

[23] O. Taghizadeh, J. Zhang, and M. Haardt, "Transmit beamforming aided amplify-and-forward MIMO full-duplex relaying with limited dynamic range". *Signal Processing,* no. 127, pp. 266-281, 2016.

[24] X. Xue, Y. Wang, X. Wang, and T. E. Bogale, "Joint source and relay precoding in multi-antenna millimeter-wave systems," *IEEE Trans. Veh. Technol.,* vol. 66, no. 6, pp. 4924–4937, Jun. 2017.

[25] S. Gong, C. Xing, Z. Fei, and S. Ma, "Millimeter-Wave secrecy beamforming designs for two-way amplify-and-forward MIMO relaying networks," *IEEE Trans. Veh. Technol.,* vol. 66, no. 3, pp. 2059–2071, Mar. 2017.

[26] Q. Shi, M. Hong, X. Fu, T.-H. Chang, "Penalty dual decomposition method for nonsmooth nonconvex optimization," [Online]. Available: https://arxiv.org/abs/1712.04767

[27] Q. Shi and M. Hong, "Penalty dual decomposition method with application in signal processing," in *Proc. Int. Conf. on Acoust. Speech Signal Process. (ICASSP),* Mar. 5–9, 2017, pp. 4059–4063.

[28] Q. Shi and M. Hong, "Spectral efficiency optimization for mmWave multiuser MIMO systems," *IEEE J. Sel. Toptics Signal Process.,* vol. 12, no. 3, pp. 455-468, Jun. 2018.

[29] Y. Cai, C. Zhao, Q. Shi, G. Y. Li, and B. Champagne, "Joint beamforming and jamming design for mmWave information surveillance systems," *IEEE J. Sel. Areas Commun.,* to appear.

[30] D. Bertsekas, *Nonlinear Programming,* 2nd ed. Belmont, MA: Athena Scientific, 1999.

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

[32] A. L. Yuille and A. Rangarajan, "The concave-convex procedure," *Neural Computation*, vol. 15, no. 4, pp. 915–936, Apr. 2003.

[33] G. R. Lanckriet and B. K. Sriperumbudur, "On the convergence of the concave-convex procedure," in *Advances in Neural Info. Processing Systems*, pp. 1759–1767, 2009.

[34] K. Shen and W. Yu, "Fractional programming for communication systemsłpart I: power control and beamforming," *IEEE Trans. Signal Process.,* vol. 66, no. 10, pp. 2616-2630, May 15, 2018.

[35] M. Ding and S. Blostein, "MIMO minimum total MSE transceiver design with imperfect CSI at both ends," *IEEE Trans. Signal Process.,* vol. 57, no. 3, pp. 1141-1150, Mar. 2009.

[36] K. Venugopal, A. Alkhateeb, N. G. Prelcic, and R. W. Heath, Jr., "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE J. Sel. Areas Commun.,* vol. 35, no. 9, pp. 1996–2009, Sep. 2017.

[37] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Areas Commun.,* vol. 8, no. 5, pp. 831–846, Oct. 2014.

[38] J. Dai, A. Liu and V. K. N. Lau, "FDD massive MIMO channel estimation with arbitrary 2D-array geometry," *IEEE Trans. Signal Process.,* to appear, 2018.

[39] L. Chen, A. Liu, X. Yuan, "Structured turbo compressed sensing for massive MIMO channel estimation using a Markov prior," *IEEE Trans. Vehi. Technol.,* vol. 67, no. 5, pp. 4635–4639, May 2018.

[40] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theorey,* vol. 46, no. 3, pp. 933-946, May 2000.

[41] M. Razaviyayn, M. Hong and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization", *SIAM Journal on Optimization.* vol. 23, no. 2, pp. 1126–1153, 2013.

[42] O. Taghizadeh, A. C. Cirik and R. Mathar, "Hardware impairments aware transceiver design for full-duplex amplify-and-forward MIMO relaying," *IEEE Trans. Wireless Commun.,* vol. 17, no. 3, pp. 1644-1659, Mar. 2018.

[43] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics in Signal Process.,* vol. 10, no. 3, pp. 501–513, April 2016.

[44] A. Alkhateeb, G. Leus and R. W. Heath, "Limited Feedback Hybrid Precoding for Multi-User Millimeter Wave Systems," *IEEE Trans. Wireless Commun.,* vol. 14, no. 11, pp. 6481-6494, Nov. 2015.

[45] A. Pizzo and L. Sanguinetti, "Optimal design of energy-efficient millimeter wave hybrid transceivers for wireless backhaul," in *Proc. International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt),* Paris, 2017, pp. 1-8.

[46] R. He, B. Ai, G. L. Stuber, G. Wang, and Z. Zhong, "Geometrical based modeling for millimeter wave MIMO mobile-to-mobile channels," *IEEE Trans. Vehi. Technol.,* vol. 67, no. 4, pp. 2848-2863, April 2018.

[47] P. F. M. Smulders and L. M. Correia, "Characterisation of propagation in 60 GHz radio channels," *Electron. Commun. Eng. J.,* vol. 9, no. 2, pp. 73-80, Apr. 1997.

[48] H. Xu, V. Kukshya, and T. S. Rappaport, "Spatial and temporal characteristics of 60-GHz indoor channels," *IEEE J. Sel. Areas Commun.,* vol. 20, no. 3, pp. 620-630, Apr. 2002.

[49] Z. Xiao, P. Xia, and X.-G. Xia, "Full-duplex millimeter-wave communication," *IEEE Wireless Commun.,* vol. 24, no. 6, pp. 136-143, Dec. 2017.

[50] H. Nicholas, *Accuracy and Stability of Numerical Algorithms*, (2nd ed.) SIAM., pp. 258, 2002.