

Robust Joint Hybrid Analog-Digital Transceiver Design for Full-Duplex mmWave Multicell Systems

Ming-Min Zhao, Yunlong Cai, Min-Jian Zhao, Ying Xu, and Lajos Hanzo

Abstract—In this work, we investigate a full-duplex (FD) millimeter wave (mmWave) multicell system, where the BS of each cell receives signals from uplink (UL) users and transmits signals to downlink (DL) users at the same time, over the same frequency band. We maximize the sum rate lower bound of the FD multicell system by jointly optimizing the digital and analog beamforming matrices at the base station (BS) and the transmit power levels of the UL users under total transmit power constraints and unit-modulus constraints (due to the analog beamforming matrices), in the presence of imperfect channel state information (CSI). The problem under study is very challenging due to the highly non-convexity of the objective function and constraints. We transform this problem into an equivalent but more tractable form and propose a novel iterative algorithm based on the penalty dual decomposition (PDD) to solve it. The proposed algorithm is guaranteed to converge to the set of Karush-Kuhn-Tucker (KKT) solutions of the original problem. Moreover, we also extend our proposed algorithm to the structure of subarray. Simulation results validate the effectiveness of the proposed algorithm as compared with conventional nonrobust and half-duplex (HD) algorithms.

Index Terms—Full-duplex, mmWave communications, multicell, robust transceiver design.

I. INTRODUCTION

Millimeter wave (mmWave) transmission with large-scale antenna arrays has been regarded as a promising solution to mitigate the severe spectrum shortage and support the ever-increasing data traffic for the next generation wireless communication systems [1]–[5]. However, the large number of antennas in such a system poses two major challenges, i.e., (1) many radio-frequency (RF) chains are needed, which increases hardware cost and power consumption; (2) acquisition of large-dimensional channel state information (CSI) is required, which often incurs severe CSI errors. One promising solution

to resolve these problems is to utilize hybrid analog-digital (AD) beamforming architectures, i.e., combining the digital beamformers in the baseband and the analog beamformers in the RF domain. As a result, the hybrid beamforming techniques have been widely investigated in the literature [6]–[8].

Some pioneering investigations have been conducted in [6], [9]–[15] for hybrid transmit and receive beamforming optimization. Specifically, the authors of [9] exploited the spatial structure of mmWave channels, the hybrid beamforming problem was formulated as a sparse reconstruction problem and a modified orthogonal matching pursuit (OMP) algorithm was developed to solve it. In [10], the authors formulated the hybrid transceiver design problem as a nonlinear least-square problem and solved it iteratively in a greedy manner, where in each step one-dimensional search with the Nelder-Mead simplex method is employed. The work of [6] treated the hybrid beamforming design as a matrix factorization problem and developed an alternating minimization algorithm based on the manifold optimization. In [11], the authors decoupled the matrix factorization problem into a series of convex sub-problems and also solved it via alternating optimization. In [12], a heuristic algorithm was proposed, where the analog beamformer is designed independently and then, given the analog precoder, the other transceiver parameters are sequentially designed. The simulation results showed that a near fully digital beamforming performance can be obtained. Besides, codebook-based hybrid beamforming designs were studied in [13], [14]. Furthermore, the studies in [7], [16]–[18] investigated the hybrid transceiver design problem in multiuser mmWave systems.

As another promising technique for the next-generation wireless communication systems, the full-duplex (FD) transmission is able to enhance the utilization of existing spectrum resources by enabling transmission and reception of user signals in the same time/frequency resource block [19]–[22]. By integrating FD with mmWave communications, it can potentially harvest the benefits of these two emerging techniques to further improve spectral efficiency [23]–[25]. Due to the presence of self-interference (SI), FD communications was conventionally believed to be infeasible. In recent years, a number of breakthroughs in hardware design have been achieved to show that SI can be canceled with an acceptable level of residual SI [26], [27], which enables the feasibility of FD transmissions. However, in order to fully exploit the potential performance gain of FD communications, the impact of the interference that is caused by FD operations should be taken into consideration. Therefore, more sophisticated

M.-M. Zhao, Y. Cai, M.-J. Zhao and Y. Xu are with the College of ISEE, Zhejiang University, Hangzhou, 310027, China (e-mail: {zmblack, ylcai, mjzhao, xuying0124}@zju.edu.cn). L. Hanzo is with the Department of ECS, University of Southampton, U.K. (Email: lh@ecs.soton.ac.uk). (*Correspondence author: Min-Jian Zhao; Yunlong Cai*)

The work of M. M. Zhao was supported in part by the National Natural Science Foundation of China under Grant 91938202, in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LQ20F010010 and in part by the Fundamental Research Funds for the Central Universities under Grant 2019QNA5011. The work of Y. Cai was supported in part by the National Natural Science Foundation of China under Grants 61831004 and 61971376, and in part by the Zhejiang Provincial Natural Science Foundation for Distinguished Young Scholars under Grant LR19F010002. The work of M. J. Zhao was supported by the National Key Research and Development Project under grant 2018YFB1802303.

L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/N004558/1, EP/P034284/1, EP/P034284/1, EP/P003990/1 (COALESCE), of the Royal Society's Global Challenges Research Fund Grant as well as of the European Research Council's Advanced Fellow Grant QuantCom.

interference management techniques need to be carefully developed. Recently, a number of studies have been proposed to handle interference and enable the FD-based mmWave communications [28]–[33]. To be specific, in [28], the authors developed adaptive self-interference cancellation and transmit power adaptation algorithms to enhance the energy efficiency of FD mmWave relaying systems. The study of [29] proposed the use of FD relays in mmWave communication for wireless backhaul links between the cells. The authors of [30] studied FD relay selection and power allocation schemes for mmWave device-to-device communications. In order to enlarge the range of mmWave transmissions, the work of [31] employed multiple FD buffered relays and analyzed the overall network performance. In [32], a quality-of-service (QoS)-aware FD concurrent scheduling algorithm was proposed to maximize the number of flows under the QoS requirements. In addition, the authors of [33] investigated the joint design of hybrid AD beamformers in FD mmWave communications.

Most of the aforementioned studies are based on perfect CSI, however, this condition can hardly be true in practice due to various factors such as estimation/quantization errors, pilot contamination, feedback delays, etc. Therefore, developing robust hybrid AD transceivers for FD mmWave systems has attracted great attention recently [34]–[36]. In [34], the authors designed the transceivers based on imperfect CSI in a mmWave amplify-and-forward (AF) relay system. In particular, they derived an approximation of the average received signal-to-noise-ratio (SNR) as the objective function, and developed an OMP-based algorithm. The work of [35] considered the imperfection of loopback self-interference knowledge and proposed a sparse approximation technique to reduce the hardware complexity. In [36], a robust joint transceiver algorithm was proposed for a FD mmWave MIMO relay system in the presence of stochastic CSI errors.

However, to the best of our knowledge, the robust designs for FD mmWave multicell systems with imperfect CSI have not been reported yet. To fill this gap, in this work, we study the robust joint hybrid transceiver design problem in such an important system with serious interference and CSI errors. Specifically, we assume that the base station (BS) of each cell in the system enables transmission and reception of user signals in the same time/frequency resource block. To proceed, we first derive the worst-case sum rate expression for the considered mmWave multicell system, and then jointly optimize the hybrid transmit and receive beamforming matrices of the BSs and the transmit power levels of the uplink (UL) users, in order to maximize the worst-case sum rate under transmit power constraints and the unit-modulus constraints on the elements of analog beamforming matrices. This problem is very challenging due to the highly non-convex objective function and constraints, and we propose to transform this problem into an equivalent but more tractable form by introducing auxiliary variables and utilizing the weighted minimum mean squared error (WMMSE) method [37]. Then, this resultant problem is tackled by a novel penalty dual decomposition (PDD)-based algorithm, which consists of two loops. The inner loop of the proposed algorithm solves the augmented Lagrangian (AL) problem of the converted problem to some accuracy via a

block successive upper-bound minimization (BSUM) method [38], while the outer loop deals with the updating of the dual variables and a certain penalty parameter. We show that closed-form solutions can be obtained in each step of the proposed algorithm and the convergence to the set of Karush-Kuhn-Tucker (KKT) solutions of the original problem can be guaranteed. Moreover, we extend the proposed algorithm to the subarray structure. Simulation results validate the effectiveness of the proposed algorithm, compared to the conventional half-duplex (HD) algorithms.

The main contributions of this work can be summarized as follows:

(1) We propose a novel FD mmWave multicell system and formulate an important hybrid transceiver design problem in the presence of imperfect CSI.

(2) To address this challenging problem, we propose a PDD-based algorithm, which is guaranteed to converge to the set of KKT solutions of the original problem. To derive the proposed algorithm, we introduce a method to handle the nontrivial unit-modulus constraints caused by the hybrid beamforming structure and to facilitate distributed updating of the optimization variables by properly introducing auxiliary variables.

(3) Extension of the proposed algorithm to the subarray structure is also provided together with a detailed complexity analysis. Furthermore, elaborated simulations are conducted to show the advantages of the proposed algorithms over the conventional schemes.

The rest of this paper is organized as follows. In Sections II, we present the investigated system model and formulate the problem of interest. In Section III, we first provide a brief introduction of the PDD optimization framework, and then introduce our PDD-based robust joint transceiver algorithm, together with detailed complexity analysis. In Section IV, we extend the proposed algorithm to the subarray structure. Section V presents numerical results to evaluate the performance of the proposed algorithm and Section VI concludes the paper.

Notations: $\mathbb{E}\{\cdot\}$ denotes the expectation operator. Scalars, vectors and matrices are respectively denoted by lower case, boldface lower case and boldface upper case letters. $\text{Tr}(\mathbf{A})$, \mathbf{A}^T , \mathbf{A}^* , \mathbf{A}^H , \mathbf{A}^{-1} and \mathbf{A}^\dagger represent the trace, transpose, conjugate, conjugate transpose, inverse and pseudo-inverse of matrix \mathbf{A} , respectively. $\mathbf{A}(a, \cdot)$ ($\mathbf{A}(\cdot, b)$) denotes a vector obtained by selecting the a -th row (b -th column) of \mathbf{A} and $\mathbf{A}(a, b)$ represents the element on the a -th row and b -th column. $\mathbf{A}_{a:b}$ denotes the collection of variables with subscripts a to b . The operators $\text{diag}\{a_1, \dots, a_M\}$ and $\text{blkdiag}\{\mathbf{a}_1, \dots, \mathbf{a}_M\}$ denote diagonal matrices with $\{a_1, \dots, a_M\}$ and $\{\mathbf{a}_1, \dots, \mathbf{a}_M\}$ denoting the diagonal elements and vectors. The operator $\text{vec}(\cdot)$ stacks the elements of a matrix in one long column vector. By \mathbf{I} and $\mathbf{0}$, we denote the identity and zero matrices with appropriate dimensions. $\|\cdot\|$, $|\cdot|$ and $\det(\cdot)$ denote the Euclidean norm of a complex vector/matrix, the absolute value of a complex scalar and the determinant of a matrix. $\mathbb{C}^{m \times n}$ ($\mathbb{R}^{m \times n}$) denotes the space of $m \times n$ complex (real) matrices. $\Re\{\cdot\}$ and $\Im\{\cdot\}$ respectively denote the real part and imaginary part of their arguments. The set difference is defined as $\mathcal{A} \setminus \mathcal{B} \triangleq \{x | x \in \mathcal{A}, x \notin \mathcal{B}\}$. The operator $\mathbb{P}_\Omega\{\mathbf{A}\} \triangleq$

$\min_{\mathbf{B} \in \Omega} \|\mathbf{A} - \mathbf{B}\|$ denotes the projection of a point \mathbf{A} onto a set Ω , if the set Ω is a sphere of radius r centered at the origin, i.e., $\Omega = \{\mathbf{A} \mid \|\mathbf{A}\| \leq r\}$, then $\mathbb{P}_\Omega\{\mathbf{A}\} = r \frac{\mathbf{A}}{\|\mathbf{A}\| + \max(0, r - \|\mathbf{A}\|)}$.

II. SYSTEM MODEL AND PROBLEM FORMULATION

In this section, we provide the system model and formulate the considered worst-case sum rate maximization problem.

A. System Model

Let us consider a FD mmWave multiuser multicell system, which consists of N cells and each cell contains one FD BS, K downlink (DL) users and J UL users, as shown in Fig. 1. The BS is equipped with N_t transmit antennas, N_r receive antennas and N_{RF} ($N_{\text{RF}} \leq \min\{N_t, N_r\}$) RF chains for the transmit and receive processing, respectively. Each user operates in the HD mode and is equipped with a single antenna. To reserve enough degrees of freedom for signal detection, we assume that $\max\{K, J\} \leq N_{\text{RF}}$ is satisfied. Furthermore, in order to reduce hardware complexity, we consider a hybrid AD beamforming architecture for the BSs [12].

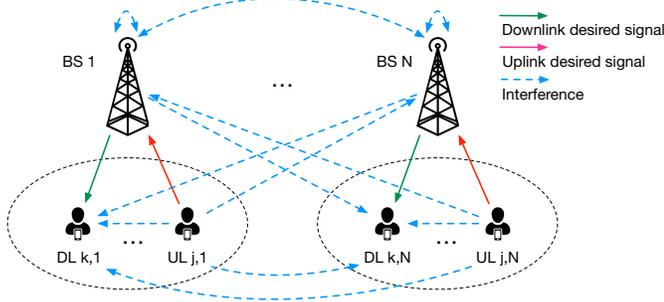


Fig. 1: Millimeter wave FD multicell system

Let $\mathbf{x}^{(n)} = [x_1^{(n)}, \dots, x_K^{(n)}]^T \in \mathbb{C}^{K \times 1}$ represent the transmit symbol vector at BS n , $n \in \mathcal{N} \triangleq \{1, \dots, N\}$, whose elements are statistically independent with zero mean and $\mathbb{E}\{|x_k^{(n)}|^2\} = 1$, $k \in \mathcal{K} \triangleq \{1, \dots, K\}$. Then, the transmit signal vector at the BS in cell n , can be written as $\mathbf{x}_B^{(n)} = \mathbf{V}_{\text{RF}}^{(n)} \mathbf{V}_{\text{BB}}^{(n)} \mathbf{x}^{(n)}$, where $\mathbf{V}_{\text{BB}}^{(n)} \in \mathbb{C}^{N_{\text{RF}} \times K}$ and $\mathbf{V}_{\text{RF}}^{(n)} \in \mathbb{C}^{N_t \times N_{\text{RF}}}$ represent the digital and analog transmit beamforming matrices at BS n , respectively. Similarly, let $\mathbf{s}^{(n)} = [s_1^{(n)}, \dots, s_J^{(n)}]^T$ represent the $J \times 1$ transmit symbol vector of the UL users in cell n , whose elements are also statistically independent with zero mean and $\mathbb{E}\{|s_j^{(n)}|^2\} = 1$, $j \in \mathcal{J} \triangleq \{1, \dots, J\}$. The transmit signal at the j -th UL user in cell n can be written as $y_j^{(n)} = \sqrt{P_j^{(n)}} s_j^{(n)}$, where $P_j^{(n)}$ denotes the transmit power of the j -th UL user.

The received signal at the k -th DL user in cell n can be written as

$$\begin{aligned} y_k^{(n)} = & \underbrace{\sum_{m \in \mathcal{N}} \mathbf{h}_k^{(n,m)H} \mathbf{V}_{\text{RF}}^{(m)} \mathbf{V}_{\text{BB}}^{(m)} \mathbf{x}^{(m)}}_{\text{received signal from the BSs}} \\ & + \underbrace{\sum_{m \in \mathcal{N}} \sum_{j \in \mathcal{J}} c_{k,j}^{(n,m)*} \sqrt{P_j^{(m)}} s_j^{(m)}}_{\text{the UL-to-DL interference}} + n_k^{(n)}, \end{aligned} \quad (1)$$

where $\mathbf{h}_k^{(n,m)} \in \mathbb{C}^{N_t \times 1}$ represents the channel vector between BS m and the k -th DL user in cell n , $c_{k,j}^{(n,m)} \in \mathbb{C}^{1 \times 1}$ denotes the channel coefficient between the j -th UL user in cell m and the k -th DL user in cell n , and $n_k^{(n)}$ is the channel noise at user k with zero mean and variance $\mathbb{E}\{|n_k^{(n)}|^2\} = \sigma_k^2$.

The received signal at BS n can be written as

$$\begin{aligned} \mathbf{y}_{\text{BS}}^{(n)} = & \sum_{m \in \mathcal{N}} \sum_{j \in \mathcal{J}} \mathbf{g}_j^{(n,m)} \sqrt{P_j^{(m)}} s_j^{(m)} \\ & + \sum_{m \in \mathcal{N}} \mathbf{F}^{(n,m)} \mathbf{V}_{\text{RF}}^{(m)} \mathbf{V}_{\text{BB}}^{(m)} \mathbf{x}^{(m)} + \mathbf{n}_b^{(n)}, \end{aligned} \quad (2)$$

where $\mathbf{g}_j^{(n,m)} \in \mathbb{C}^{N_r \times 1}$ denotes the channel vector between the j -th UL user in cell m and BS n , $\mathbf{F}^{(n,m)} \in \mathbb{C}^{N_r \times N_t}$ is the channel matrix between BS m and BS n , and $\mathbf{n}_b^{(n)}$ represents the complex circular Gaussian noise vector at BS n with zero mean and covariance $\mathbb{E}\{\mathbf{n}_b^{(n)} \mathbf{n}_b^{(n)H}\} = \sigma_{\text{BS}}^2 \mathbf{I}$. Based on certain interference cancellation techniques, we assume that the intracell self-interference of the BSs can be eliminated [39]. As a result, the processed signal at BS n can be written as

$$\begin{aligned} \mathbf{r}_{\text{BS}}^{(n)} = & \mathbf{U}_{\text{BB}}^{(n)H} \mathbf{U}_{\text{RF}}^{(n)H} \left(\underbrace{\sum_{m \in \mathcal{N}} \mathbf{G}^{(n,m)} \sqrt{\mathbf{P}^{(m)}} \mathbf{s}^{(m)}}_{\text{received signal from the UL users}} \right. \\ & \left. + \underbrace{\sum_{m \in \mathcal{N} \setminus n} \mathbf{F}^{(n,m)} \mathbf{V}_{\text{RF}}^{(m)} \mathbf{V}_{\text{BB}}^{(m)} \mathbf{x}^{(m)} + \mathbf{n}_b^{(n)}}_{\text{the DL-to-UL interference}} \right), \end{aligned} \quad (3)$$

where $\mathbf{U}_{\text{RF}}^{(n)} \in \mathbb{C}^{N_r \times N_{\text{RF}}}$ represents the analog receive beamforming matrix at BS n , $\mathbf{U}_{\text{BB}}^{(n)} \in \mathbb{C}^{N_{\text{RF}} \times J}$ represents the digital receive beamforming matrix, and we have $\mathbf{G}^{(n,m)} = [\mathbf{g}_1^{(n,m)}, \dots, \mathbf{g}_J^{(n,m)}]$ and $\mathbf{P}^{(m)} = \text{diag}\{P_1^{(m)}, \dots, P_J^{(m)}\}$.

By taking the CSI errors into consideration, the channel matrix between BS m and the DL users in cell n , i.e., $\mathbf{H}^{(n,m)} = [\mathbf{h}_1^{(n,m)}, \dots, \mathbf{h}_K^{(n,m)}]$, can be expressed as

$$\mathbf{H}^{(n,m)} = \hat{\mathbf{H}}^{(n,m)} + \Delta \mathbf{H}^{(n,m)}, \quad (4)$$

where $\hat{\mathbf{H}}^{(n,m)} = [\hat{\mathbf{h}}_1^{(n,m)}, \dots, \hat{\mathbf{h}}_K^{(n,m)}]^H$ denotes the estimated channel matrix and $\Delta \mathbf{H}^{(n,m)} = [\Delta \mathbf{h}_1^{(n,m)}, \dots, \Delta \mathbf{h}_K^{(n,m)}]^H$ represents the corresponding channel error matrix.¹ In this work, $\Delta \mathbf{H}^{(n,m)}$ is assumed to be independent of $\hat{\mathbf{H}}^{(n,m)}$ and satisfies $\Delta \mathbf{H}^{(n,m)} \sim \mathcal{CN}(\mathbf{0}, \sigma_h^2 \mathbf{I})$, where σ_h^2 denotes the variance of CSI errors. Similarly, the channel matrices between the BSs $\{\mathbf{F}^{(n,m)}\}$, the channel matrices between the BSs and the UL users $\{\mathbf{G}^{(n,m)}\}$, as well as the channel coefficients between the UL users and the DL users $\{c_{k,j}^{(n,m)}\}$ can be modeled as follows:

$$\mathbf{F}^{(n,m)} = \hat{\mathbf{F}}^{(n,m)} + \Delta \mathbf{F}^{(n,m)}, \quad \Delta \mathbf{F}^{(n,m)} \sim \mathcal{CN}(\mathbf{0}, \sigma_f^2 \mathbf{I}), \quad (5)$$

$$\mathbf{G}^{(n,m)} = \hat{\mathbf{G}}^{(n,m)} + \Delta \mathbf{G}^{(n,m)}, \quad \Delta \mathbf{G}^{(n,m)} \sim \mathcal{CN}(\mathbf{0}, \sigma_g^2 \mathbf{I}), \quad (6)$$

¹In practice, the CSI of mmWave communication systems can be obtained by employing certain compressed sensing channel estimation algorithms, such as those in [40], [41], and the resulting CSI errors obey a multivariate Gaussian distribution under the case of additional Gaussian noise.

$$c_{k,j}^{(n,m)} = \hat{c}_{k,j}^{(n,m)} + \Delta c_{k,j}^{(n,m)}, \Delta c_{k,j}^{(n,m)} \sim \mathcal{CN}(0, \sigma_c^2). \quad (7)$$

Furthermore, in this work, we primarily focus on the so called *coordinated beamforming* scheme, where a central processing (CP) node collects all the estimated CSI, computes the required variables and then distributes them to the respective nodes. This is accomplished by exchanging the CSI information between the nodes through backhaul links [21], [42], [43].

B. Problem Formulation

Let us first derive the deterministic lower bound (DLB) of the sum rate at the user side. By expanding the mutual information [44], [45] of user k into differential entropy, we obtain

$$\mathcal{I}(x_k^{(n)}; y_k^{(n)} | \hat{\mathbf{H}}^{(n)}, \hat{\mathbf{C}}^{(n)}) = H(x_k^{(n)} | \hat{\mathbf{H}}^{(n)}, \hat{\mathbf{C}}^{(n)}) - H(x_k^{(n)} | y_k^{(n)}, \hat{\mathbf{H}}^{(n)}, \hat{\mathbf{C}}^{(n)}), \quad (8)$$

where $\hat{\mathbf{H}}^{(n)} \triangleq \{\hat{\mathbf{H}}^{(n,1)}, \dots, \hat{\mathbf{H}}^{(n,N)}\}$ and $\hat{\mathbf{C}}^{(n)} \triangleq \{\hat{c}_{k,j}^{(n,m)}\}_{\forall k,j,m}$. It can be seen that $H(x_k^{(n)} | \hat{\mathbf{H}}^{(n)}, \hat{\mathbf{C}}^{(n)})$ becomes $\log(\pi e)$ [46], in order to obtain an upper bound for $H(x_k^{(n)} | y_k^{(n)}, \hat{\mathbf{H}}^{(n)}, \hat{\mathbf{C}}^{(n)})$, we can rewrite (1) as follows:

$$\begin{aligned} y_k^{(n)} &= \sum_{m \in \mathcal{N}} \hat{\mathbf{h}}_k^{(n,m)H} \mathbf{V}_{\text{RF}}^{(m)} \mathbf{V}_{\text{BB}}^{(m)} \mathbf{x}^{(m)} \\ &+ \sum_{m \in \mathcal{N}} \sum_{j \in \mathcal{J}} \hat{c}_{k,j}^{(n,m)*} \sqrt{P_j^{(m)}} s_j^{(m)} \\ &+ \sum_{m \in \mathcal{N}} \Delta \mathbf{h}_k^{(n,m)H} \mathbf{V}_{\text{RF}}^{(m)} \mathbf{V}_{\text{BB}}^{(m)} \mathbf{x}^{(m)} \\ &+ \sum_{m \in \mathcal{N}} \sum_{j \in \mathcal{J}} \Delta c_{k,j}^{(n,m)*} \sqrt{P_j^{(m)}} s_j^{(m)} + n_k^{(n)}. \end{aligned} \quad (9)$$

Using the fact that the entropy of a random variable with given variance is upper-bounded by the entropy of a Gaussian random variable with the same variance [46], $H(x_k^{(n)} | y_k^{(n)}, \hat{\mathbf{H}}^{(n)}, \hat{\mathbf{C}}^{(n)})$ can be upper bounded by the entropy of a Gaussian random variable with variance being $\mathbb{E}_{\{x_k^{(n)}, n_k^{(n)}, \Delta \mathbf{H}^{(n,m)}, \Delta c_{k,j}^{(n,m)}\}}((x_k^{(n)} - u_k^{(n)*} y_k^{(n)})(x_k^{(n)} - u_k^{(n)*} y_k^{(n)})^* | \hat{\mathbf{H}}^{(n)}, \hat{\mathbf{C}}^{(n)})$, where $u_k^{(n)*} y_k^{(n)}$ is the (linear) MMSE estimation of $x_k^{(n)}$. Then, it can be observed that the following inequality holds:

$$\begin{aligned} &H(x_k^{(n)} | y_k^{(n)}, \hat{\mathbf{H}}, \hat{\mathbf{C}}) \\ &\leq \log \left(\pi e \left(1 - \frac{|\hat{\mathbf{h}}_k^{(n,n)H} \mathbf{V}_{\text{RF}}^{(n)} \mathbf{V}_{\text{BB},k}^{(n)}|^2}{|\hat{\mathbf{h}}_k^{(n,n)H} \mathbf{V}_{\text{RF}}^{(n)} \mathbf{V}_{\text{BB},k}^{(n)}|^2 + \Upsilon_k^{(n)}} \right) \right), \end{aligned} \quad (10)$$

where

$$\begin{aligned} \Upsilon_k^{(n)} &\triangleq \sum_{m \in \mathcal{N} \setminus n} \sum_{i \in \mathcal{K}} |\hat{\mathbf{h}}_k^{(n,m)H} \mathbf{V}_{\text{RF}}^{(m)} \mathbf{V}_{\text{BB},i}^{(m)}|^2 \\ &+ \sum_{m \in \mathcal{N}} \sum_{i \in \mathcal{K}} \sigma_h^2 \text{Tr}(\mathbf{V}_{\text{RF}}^{(m)} \mathbf{V}_{\text{BB},i}^{(m)} \mathbf{V}_{\text{BB},i}^{(m)H} \mathbf{V}_{\text{RF}}^{(m)H}) \\ &+ \sum_{k' \in \mathcal{K} \setminus k} |\hat{\mathbf{h}}_k^{(n,n)H} \mathbf{V}_{\text{RF}}^{(n)} \mathbf{V}_{\text{BB},k'}^{(n)}|^2 \\ &+ \underbrace{\sum_{m \in \mathcal{N}} \sum_{j \in \mathcal{J}} |\hat{c}_{k,j}^{(n,m)}|^2 P_j^{(m)} + \sum_{m \in \mathcal{N}} \sum_{j \in \mathcal{J}} \sigma_c^2 P_j^{(m)} + \sigma_c^2}_{\text{average UL-to-DL interference power}}. \end{aligned} \quad (11)$$

Hence, by plugging (10) into (8), we can obtain the DLB of mutual information for user k as follows:

$$\begin{aligned} \mathcal{I}(x_k^{(n)}; y_k^{(n)} | \hat{\mathbf{H}}, \hat{\mathbf{C}}) &\geq \tilde{R}_{\text{DLB}}^{(n,k)}(\{\mathbf{V}_{\text{RF}}^{(n)}, \mathbf{V}_{\text{BB}}^{(n)}, \mathbf{P}^{(n)}\}) \\ &\triangleq \log \left(1 + \frac{|\hat{\mathbf{h}}_k^{(n,n)H} \mathbf{V}_{\text{RF}}^{(n)} \mathbf{V}_{\text{BB},k}^{(n)}|^2}{\Upsilon_k^{(n)}} \right). \end{aligned} \quad (12)$$

By following a similar procedure, we can have the DLB of the sum rate at BS n , which can be expressed as

$$\begin{aligned} &\bar{R}_{\text{DLB}}^{(n)}(\{\mathbf{U}_{\text{RF}}^{(n)}, \mathbf{P}^{(n)}\}) \\ &\triangleq \log \det \left(\mathbf{I} + \mathbf{U}_{\text{RF}}^{(n)H} \hat{\mathbf{G}}^{(n,n)} \mathbf{P}^{(n)} \hat{\mathbf{G}}^{(n,n)H} \mathbf{U}_{\text{RF}}^{(n)} (\mathbf{\Upsilon}_{\text{BS}}^{(n)})^{-1} \right), \end{aligned} \quad (13)$$

where

$$\begin{aligned} \mathbf{\Upsilon}_{\text{BS}}^{(n)} &\triangleq \mathbf{U}_{\text{RF}}^{(n)H} \left(\sum_{m \in \mathcal{N} \setminus n} \hat{\mathbf{G}}^{(n,m)} \mathbf{P}^{(m)} \hat{\mathbf{G}}^{(n,m)H} \right. \\ &+ \underbrace{\sum_{m \in \mathcal{N} \setminus n} \hat{\mathbf{F}}^{(n,m)} \mathbf{V}_{\text{RF}}^{(m)} \mathbf{V}_{\text{BB}}^{(m)} \mathbf{V}_{\text{BB}}^{(m)H} \mathbf{V}_{\text{RF}}^{(m)H} \hat{\mathbf{F}}^{(n,m)H}}_{(a)} \\ &+ \underbrace{\sum_{m \in \mathcal{N} \setminus n} \sigma_f^2 \text{Tr}(\mathbf{V}_{\text{RF}}^{(m)} \mathbf{V}_{\text{BB}}^{(m)} \mathbf{V}_{\text{BB}}^{(m)H} \mathbf{V}_{\text{RF}}^{(m)H}) \mathbf{I}}_{(b)} \\ &\left. + \sum_{m \in \mathcal{N}} \sigma_g^2 \text{Tr}(\mathbf{P}^{(m)}) \mathbf{I} + \sigma_{\text{BS}}^2 \mathbf{I} \right) \mathbf{U}_{\text{RF}}^{(n)}, \end{aligned} \quad (14)$$

and (a)+(b) denotes the average DL-to-UL interference power.

In this work, we aim to maximize the sum of the lower bounds of the mutual information over all users and cells by optimizing the hybrid AD beamforming matrices at the BSs and the transmit power levels of the UL users, where the optimization problem can be formulated as follows:

$$\max_{\{\mathbf{x}^{(n)} \setminus \mathbf{U}_{\text{BB}}^{(n)}\}} \sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} \tilde{R}_{\text{DLB}}^{(n,k)} + \sum_{n \in \mathcal{N}} \bar{R}_{\text{DLB}}^{(n)} \quad (15a)$$

$$\text{s.t. } P_j^{(n)} \leq P_u, \forall j, n, \quad (15b)$$

$$\|\mathbf{V}_{\text{RF}}^{(n)} \mathbf{V}_{\text{BB}}^{(n)}\|^2 \leq P_B, \forall n, \quad (15c)$$

$$|\mathbf{V}_{\text{RF}}^{(n)}(i, j)| = 1, |\mathbf{U}_{\text{RF}}^{(n)}(i, j)| = 1, \forall i, j, n, \quad (15d)$$

where $\mathbf{x}^{(n)} \triangleq \{\mathbf{V}_{\text{RF}}^{(n)}, \mathbf{V}_{\text{BB}}^{(n)}, \mathbf{U}_{\text{RF}}^{(n)}, \mathbf{U}_{\text{BB}}^{(n)}, \mathbf{P}^{(n)}\}$, (15b) and (15c) denote the power constraints of the UL users and the BSs, and P_u and P_B denote the power budgets. (15d) are unit-modulus constraints that are due to the practical implementations of the analog transmit and receive beamformers, i.e., using low-cost phase-shifters. Note that problem (15) is extremely difficult to solve due to the coupling of the AD beamformers, the unit-modulus constraints and the highly non-convex objective function. In the following, by utilizing the PDD optimization framework [47]–[49], we propose an efficient robust joint transceiver algorithm to address this problem.

III. PROPOSED ROBUST JOINT DESIGN BASED ON THE PDD FRAMEWORK

In this section, we propose to develop a PDD-based robust hybrid AD transceiver algorithm based on solving problem (15). To tackle the difficulty arising from the non-convex objective function, i.e., the DLBs of mutual information for the

users and BSs, containing the $\log(\cdot)$ and $\log \det(\cdot)$ terms, we first transform it into an equivalent but simpler form and then handle the resultant problem with the aid of the PDD method. Note that the PDD method is a recently proposed optimization framework [47], in which particular designs need to be developed for a given problem. Hence, our robust transceiver algorithm based on the PDD framework is nontrivial.

A. A Brief Introduction to the PDD Optimization Framework

In this subsection, we introduce the PDD framework in order to expose its intrinsic idea. In general, the PDD method can be applied to solve a family of non-convex non-smooth optimization problems subject to non-convex coupling constraints. More specifically, consider the following block-structured optimization problem:

$$\begin{aligned} \min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) \\ \text{s.t. } \mathbf{h}(\mathbf{x}) = \mathbf{0}, \\ \mathbf{g}_i(\mathbf{x}_i) \leq \mathbf{0}, \quad i = \{1, 2, \dots, q\}, \end{aligned} \quad (16)$$

where the optimization variable $\mathbf{x} \in \mathbb{R}^Q$ can be decomposed as $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_q)$, the feasible set \mathcal{X} is the Cartesian product of q closed convex sets, i.e., $\mathcal{X} \triangleq \prod_{i=1}^q \mathcal{X}_i$ with $\mathcal{X}_i \in \mathbb{R}^{q_i \times 1}$, $\sum_{i=1}^q q_i = Q$ and $\mathbf{x}_i \in \mathcal{X}_i$. The objective function $f(\mathbf{x})$ is a scalar continuously differentiable function, $\mathbf{h}(\mathbf{x}) \in \mathbb{R}^{p \times 1}$ is a vector of p continuously differentiable functions and $\mathbf{g}_i(\mathbf{x}_i) \in \mathbb{R}^{q_i \times 1}$ is a vector of q_i continuously differentiable functions. The detailed steps of the PDD method are presented in **Algorithm 1**, where Step 3 employs the BSUM method to update the primal variables, the control parameter η^r determines how often the AL method and the penalty method are carried out, the parameter $\epsilon^r > 0$ controls the accuracy of the BSUM method, ρ^r denotes the penalty parameter and c is a fraction which should be appropriately chosen to control the decreasing speed of the penalty parameter. The BSUM method invokes an iterative process to solve the AL problem $\mathcal{P}(\rho^r, \boldsymbol{\lambda}^r)$ to some accuracy ϵ^r starting from \mathbf{x}^0 . $\mathcal{P}(\rho^r, \boldsymbol{\lambda}^r)$ is defined as follows:

$$\min_{\{\mathbf{x}_i \in \mathcal{X}_i\}} \left\{ \mathcal{L}^r(\mathbf{x}) \triangleq f(\mathbf{x}) + (\boldsymbol{\lambda}^r)^T \mathbf{h}(\mathbf{x}) + \frac{1}{2\rho^r} \|\mathbf{h}(\mathbf{x})\|^2 \right\}, \quad (17)$$

where $\tilde{\mathcal{X}}_i \triangleq \{\mathbf{x}_i | \mathbf{g}_i(\mathbf{x}_i) \leq \mathbf{0}, \mathbf{x}_i \in \mathcal{X}_i\}$. As can be seen, the inner loop of the PDD method aims to optimize the non-convex AL problem via the BSUM method, while in the outer loop, the dual variables and a penalty parameter are updated. Furthermore, if $\mathcal{P}(\rho^r, \boldsymbol{\lambda}^r)$ can be solved to some accuracy ϵ^r with $\epsilon^r \rightarrow 0$ as $r \rightarrow \infty$, it can be proved that the limit point of the PDD algorithm is a stationary solution to problem (16) under a suitable constraint qualification condition. Interested readers may refer to [47], [50] for further details.

B. Problem Transformation

In order to transform problem (15), we first resort to the WMMSE method [37] and it can be shown that maximizing

Algorithm 1 PDD-based Algorithm for Problem (16)

```

1: Initialize  $\mathbf{x}^0, \rho^0, \boldsymbol{\lambda}^0$ , set  $0 < c < 1, r = 1$ .
2: repeat
3:    $\mathbf{x}^r = \text{optimize}(\mathcal{P}(\rho^r, \boldsymbol{\lambda}^r), \mathbf{x}^{r-1}, \epsilon^r)$ ;
4:   if  $\|\mathbf{h}(\mathbf{x}^r)\|_\infty \leq \eta^r$  then
5:      $\lambda^{r+1} = \lambda^r + \frac{1}{\rho^r} \mathbf{h}(\mathbf{x}^r), \rho^{r+1} = \rho^r$ ,
6:   else
7:      $\lambda^{r+1} = \lambda^r, \rho^{r+1} = c\rho^r$ ,
8:   end if
9:    $r = r + 1$ 
10: until some termination criterion is met.

```

$\tilde{R}_{\text{DLB}}^{(n,k)}$ and $\bar{R}_{\text{DLB}}^{(n)}$ is equivalent to minimizing the following weighted mean squared errors (MSEs):

$$h^{(n,k)}(u_k^{(n)}, w_k^{(n)}, \boldsymbol{\chi}^{(n)}) \triangleq w_k^{(n)} e_k^{(n)} - \log(w_k^{(n)}), \quad (18)$$

$$g^{(n)}(\mathbf{W}_b^{(n)}, \boldsymbol{\chi}^{(n)}) \triangleq \text{Tr}(\mathbf{W}_b^{(n)} \mathbf{E}_b^{(n)}) - \log \det(\mathbf{W}_b^{(n)}), \quad (19)$$

where $u_k^{(n)}$ and $w_k^{(n)} > 0$ denote the virtual receive coefficient and the weight factor for the k -th DL user in cell n , respectively, and $\mathbf{W}_b^{(n)} \in \mathbb{C}^{J \times J}$ is the weight matrix for BS n . Note that $\{u_k^{(n)}\}$, $\{w_k^{(n)}\}$ and $\{\mathbf{W}_b^{(n)}\}$ are introduced auxiliary variables. $e_k^{(n)}$ and $\mathbf{E}_b^{(n)}$ represent the MSEs of the k -th DL user in cell n and BS n , respectively, which are given by

$$\begin{aligned} e_k^{(n)} &\triangleq \mathbb{E} \left[|u_k^{(n)} y_k^{(n)} - s_k^{(n)}|^2 \right] \\ &= |u_k^{(n)}|^2 \left(\sum_{m \in \mathcal{N}} \|\hat{\mathbf{h}}_k^{(n,m)H} \mathbf{V}_{\text{RF}}^{(m)} \mathbf{V}_{\text{BB}}^{(m)}\|^2 \right. \\ &\quad \left. + \sum_{m \in \mathcal{N}} \sum_{j \in \mathcal{K}} |\hat{c}_{k,j}^{(n,m)}|^2 P_j^{(m)} + \sum_{m \in \mathcal{N}} \sigma_h^2 \|\mathbf{V}_{\text{RF}}^{(m)} \mathbf{V}_{\text{BB}}^{(m)}\|^2 \right) \\ &\quad \left. + \sum_{m \in \mathcal{N}} \sum_{j \in \mathcal{K}} \sigma_c^2 P_j^{(m)} + \sigma_k^2 \right) \\ &\quad - 2\Re \left(u_k^{(n)*} \hat{\mathbf{h}}_k^{(n,n)H} \mathbf{V}_{\text{RF}}^{(n)} \mathbf{V}_{\text{BB}}^{(n)} \mathbf{a}_k \right) + 1 \end{aligned} \quad (20)$$

and (21) (shown at the top of the next page), where $\mathbf{a}_k = [\mathbf{0}_{k \times 1}, 1, \mathbf{0}_{(K-k) \times 1}]^T$ represents a selection vector. Then, we have the following proposition:

Proposition III.1. Problem (15) is equivalent to

$$\begin{aligned} \min_{\{\boldsymbol{\chi}^{(n)}, u_k^{(n)}, w_k^{(n)}, \mathbf{W}_b^{(n)}\}} &\sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} h^{(n,k)}(u_k^{(n)}, w_k^{(n)}, \boldsymbol{\chi}^{(n)}) \\ &+ \sum_{n \in \mathcal{N}} g^{(n)}(\mathbf{W}_b^{(n)}, \boldsymbol{\chi}^{(n)}) \end{aligned} \quad (22)$$

s.t. (15b), (15c) and (15d),

in the sense that they have the same global optimal solution. $h^{(n,k)}(u_k^{(n)}, w_k^{(n)}, \boldsymbol{\chi}^{(n)})$ is a convex function for each set of the variables $\mathbf{V}_{\text{RF}}^{(n)}, \mathbf{V}_{\text{BB}}^{(n)}, \mathbf{P}^{(n)}, u_k^{(n)}$ and $w_k^{(n)}$ when the others are given, while $g^{(n)}(\mathbf{W}_b^{(n)}, \boldsymbol{\chi}^{(n)})$ is a convex function for $\mathbf{V}_{\text{RF}}^{(n)}, \mathbf{V}_{\text{BB}}^{(n)}, \mathbf{U}_{\text{RF}}^{(n)}, \mathbf{U}_{\text{BB}}^{(n)}, \mathbf{P}^{(n)}$ and $\mathbf{W}_b^{(n)}$, respectively. Moreover, if $\{\boldsymbol{\chi}^{(n)}, u_k^{(n)}, w_k^{(n)}, \mathbf{W}_b^{(n)}\}$ is a KKT point of problem (22), then $\{\boldsymbol{\chi}^{(n)}\}$ is a KKT point of problem (15).

Proof. Please refer to Appendix A. \square

Since the terms of $\log \det(\cdot)$ and $\log(\cdot)$ are replaced by the MSE functions, which are convex in each block of variables, the investigated problem becomes more tractable. Moreover, according to Proposition III.1, the KKT solutions of problem (15) can be obtained by resorting to problem (22). However,

$$\begin{aligned}
\mathbf{E}_b^{(n)} &\triangleq \mathbb{E} \left[(\mathbf{U}_{\text{BB}}^{(n)H} \mathbf{U}_{\text{RF}}^{(n)H} \mathbf{y}_{\text{BS}}^{(n)} - \mathbf{s}^{(n)}) (\mathbf{U}_{\text{BB}}^{(n)H} \mathbf{U}_{\text{RF}}^{(n)H} \mathbf{y}_{\text{BS}}^{(n)} - \mathbf{s}^{(n)})^H \right] \\
&= \mathbf{U}_{\text{BB}}^{(n)H} \mathbf{U}_{\text{RF}}^{(n)H} \left(\sum_{m \in \mathcal{N}} \hat{\mathbf{G}}^{(n,m)} \mathbf{P}^{(m)} \hat{\mathbf{G}}^{(n,m)H} + \sum_{m \in \mathcal{N} \setminus n} \sigma_f^2 \text{Tr}(\mathbf{V}_{\text{RF}}^{(m)} \mathbf{V}_{\text{BB}}^{(m)} \mathbf{V}_{\text{BB}}^{(m)H} \mathbf{V}_{\text{RF}}^{(m)H}) \mathbf{I} \right) \\
&\quad + \sum_{m \in \mathcal{N} \setminus n} \hat{\mathbf{F}}^{(n,m)} \mathbf{V}_{\text{RF}}^{(m)} \mathbf{V}_{\text{BB}}^{(m)} \mathbf{V}_{\text{BB}}^{(m)H} \mathbf{V}_{\text{RF}}^{(m)H} \hat{\mathbf{F}}^{(n,m)H} + \sum_{m \in \mathcal{N}} \sigma_g^2 \text{Tr}(\mathbf{P}^{(m)}) \mathbf{I} + \sigma_{\text{BS}}^2 \mathbf{I} \mathbf{U}_{\text{RF}}^{(n)} \mathbf{U}_{\text{BB}}^{(n)} \\
&\quad - \mathbf{U}_{\text{BB}}^{(n)H} \mathbf{U}_{\text{RF}}^{(n)H} \mathbf{G}^{(n,n)} \mathbf{P}^{(n)\frac{1}{2}} - \mathbf{P}^{(n)\frac{1}{2}} \mathbf{G}^{(n,n)H} \mathbf{U}_{\text{RF}}^{(n)} \mathbf{U}_{\text{BB}}^{(n)} + \mathbf{I},
\end{aligned} \tag{21}$$

it is also important to mention that the variable coupling issue in the objective function and constraints of problem (22) still remains. In order to decompose problem (22) and cope with the coupling issue, we propose to introduce a set of auxiliary variables $\mathbf{V}^{(n)}$, $\mathbf{U}^{(n)}$, $\tilde{\mathbf{V}}^{(n)}$, $\tilde{\mathbf{V}}_{\text{RF}}^{(n)}$ and $\tilde{\mathbf{U}}_{\text{RF}}^{(n)}$ which satisfy $\mathbf{V}^{(n)} = \tilde{\mathbf{V}}_{\text{RF}}^{(n)} \mathbf{V}_{\text{BB}}^{(n)}$, $\mathbf{U}^{(n)} = \tilde{\mathbf{U}}_{\text{RF}}^{(n)} \mathbf{U}_{\text{BB}}^{(n)}$, $\tilde{\mathbf{V}}^{(n)} = \mathbf{V}^{(n)}$, $\mathbf{V}_{\text{RF}}^{(n)} = \tilde{\mathbf{V}}_{\text{RF}}^{(n)}$ and $\mathbf{U}_{\text{RF}}^{(n)} = \tilde{\mathbf{U}}_{\text{RF}}^{(n)}$, respectively, then problem (22) can be equivalently converted to

$$\begin{aligned}
\min_{\{\varphi^{(n)}, u_k^{(n)}, w_k^{(n)}, \mathbf{W}_b^{(n)}\}} &\sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} \tilde{h}^{(n,k)}(u_k^{(n)}, w_k^{(n)}, \varphi^{(n)}) \\
&+ \sum_{n \in \mathcal{N}} \tilde{g}^{(n)}(\mathbf{W}_b^{(n)}, \varphi^{(n)})
\end{aligned} \tag{23a}$$

$$\text{s.t. } P_j^{(n)} \leq P_u, \forall j, n, \tag{23b}$$

$$\|\tilde{\mathbf{V}}^{(n)}\|^2 \leq P_B, \forall n, \tag{23c}$$

$$\mathbf{V}^{(n)} = \tilde{\mathbf{V}}_{\text{RF}}^{(n)} \mathbf{V}_{\text{BB}}^{(n)}, \mathbf{U}^{(n)} = \tilde{\mathbf{U}}_{\text{RF}}^{(n)} \mathbf{U}_{\text{BB}}^{(n)},$$

$$\mathbf{V}^{(n)} = \tilde{\mathbf{V}}^{(n)}, \mathbf{V}_{\text{RF}}^{(n)} = \tilde{\mathbf{V}}_{\text{RF}}^{(n)}, \mathbf{U}_{\text{RF}}^{(n)} = \tilde{\mathbf{U}}_{\text{RF}}^{(n)}, \forall n, \tag{23d}$$

$$|\mathbf{V}_{\text{RF}}^{(n)}(i, j)| = 1, |\mathbf{U}_{\text{RF}}^{(n)}(i, j)| = 1, \forall i, j, n, \tag{23e}$$

where $\varphi^{(n)} \triangleq \{\chi^{(n)}, \mathbf{V}^{(n)}, \mathbf{U}^{(n)}, \tilde{\mathbf{V}}^{(n)}, \tilde{\mathbf{V}}_{\text{RF}}^{(n)}, \tilde{\mathbf{U}}_{\text{RF}}^{(n)}\}$, $\tilde{h}^{(n,k)}(u_k^{(n)}, w_k^{(n)}, \varphi^{(n)})$ and $\tilde{g}^{(n)}(\mathbf{W}_b^{(n)}, \varphi^{(n)})$ are obtained by replacing the terms in $h^{(n,k)}(u_k^{(n)}, w_k^{(n)}, \chi^{(n)})$ and $g^{(n)}(\mathbf{W}_b^{(n)}, \chi^{(n)})$ involving $\mathbf{V}_{\text{RF}}^{(n)} \mathbf{V}_{\text{BB}}^{(n)}$ and $\mathbf{U}_{\text{RF}}^{(n)} \mathbf{U}_{\text{BB}}^{(n)}$ with $\mathbf{V}^{(n)}$ and $\mathbf{U}^{(n)}$, respectively. Note that in (23d), the auxiliary variables $\{\mathbf{V}^{(n)}\}$ and $\{\mathbf{U}^{(n)}\}$ denote the hybrid transmit and receive beamforming matrices, and $\{\tilde{\mathbf{V}}^{(n)}, \tilde{\mathbf{V}}_{\text{RF}}^{(n)}, \tilde{\mathbf{U}}_{\text{RF}}^{(n)}\}$ represent the introduced redundancy copies of the variables $\{\mathbf{V}^{(n)}, \mathbf{V}_{\text{RF}}^{(n)}, \mathbf{U}_{\text{RF}}^{(n)}\}$. The purposes of introducing these auxiliary variables are two folds: 1) to resolve the coupling issue such that the resulting constraints do not contain coupling terms (except for the introduced equality constraints (23d)); 2) to decompose the optimization variables into different blocks such that the variables which appear in the same constraint can be jointly optimized in one block and the resulting subproblems can be easily solved in parallel. The necessity and benefits of these auxiliary variables will be introduced more clearly in the next subsection. To this end, it can be seen that problem (23) is now in the form of problem (16). Hence, in the following, we address problem (23) by designing the PDD-based algorithm.

C. PDD-based Algorithm for Solving Problem (23)

It is important to mention that the key and also the main difficulty in the PDD-based algorithm is to develop an efficient BSUM method (in our case, it reduces to a block coordinate

descent (BCD) algorithm since no approximation is needed) for solving the AL problem in the inner loop. Hence, the main efforts of this subsection are devoted to developing a BCD method to efficiently solve the AL problem associated with problem (23), which is defined as

$$\begin{aligned}
\min_{\{\varphi^{(n)}, u_k^{(n)}, w_k^{(n)}, \mathbf{W}_b^{(n)}\}} &\sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} \tilde{h}^{(n,k)}(u_k^{(n)}, w_k^{(n)}, \varphi^{(n)}) \\
&+ \sum_{n \in \mathcal{N}} \tilde{g}^{(n)}(\mathbf{W}_b^{(n)}, \varphi^{(n)}) \\
&+ \frac{1}{2\rho} \sum_{n \in \mathcal{N}} \left(+ \|\mathbf{V}^{(n)} - \tilde{\mathbf{V}}_{\text{RF}}^{(n)} \mathbf{V}_{\text{BB}}^{(n)} + \rho \lambda_2^{(n)}\|^2 \right. \\
&+ \|\mathbf{U}^{(n)} - \tilde{\mathbf{U}}_{\text{RF}}^{(n)} \mathbf{U}_{\text{BB}}^{(n)} + \rho \lambda_3^{(n)}\|^2 + \|\mathbf{V}^{(n)} - \tilde{\mathbf{V}}^{(n)} + \rho \lambda_1^{(n)}\|^2 \\
&+ \|\mathbf{V}_{\text{RF}}^{(n)} - \tilde{\mathbf{V}}_{\text{RF}}^{(n)} + \rho \lambda_4^{(n)}\|^2 + \|\mathbf{U}_{\text{RF}}^{(n)} - \tilde{\mathbf{U}}_{\text{RF}}^{(n)} + \rho \lambda_5^{(n)}\|^2 \left. \right)
\end{aligned} \tag{24}$$

s.t. (23b), (23c) and (23e),

where $\{\lambda_i^{(n)}\}$, $i \in \{1, \dots, 5\}$, denote the Lagrange multipliers associated with constraint (23d), and ρ denotes a penalty parameter. In problem (24), we have dualized and penalized the equality constraints in (23d), and then added some constant terms to obtain the square terms. Note that the added constants will not effect the optimality of the AL problem. By transforming problem (23) into problem (24), we are able to effectively tackle the equality constraints in (23d) by properly adjusting ρ and $\{\lambda_i^{(n)}\}$, while the other constraints, i.e., (23b), (23c) and (23e), can be addressed by directly optimizing problem (24). It can be readily seen that the constraints of problem (24) are separable among the design variables. Hence, we propose to use a BCD-type algorithm to solve problem (24), where the original optimization variables are divided into the following three blocks: $\mathcal{B}_1 = \{u_k^{(n)}, \mathbf{U}^{(n)}, \mathbf{V}_{\text{BB}}^{(n)}, \mathbf{V}_{\text{RF}}^{(n)}, \tilde{\mathbf{V}}^{(n)}\}$, $\mathcal{B}_2 = \{\mathbf{V}^{(n)}, \mathbf{P}^{(n)}, \mathbf{U}_{\text{BB}}^{(n)}, \mathbf{U}_{\text{RF}}^{(n)}\}$ and $\mathcal{B}_3 = \{w_k^{(n)}, \tilde{\mathbf{U}}_{\text{RF}}^{(n)}, \tilde{\mathbf{V}}_{\text{RF}}^{(n)}, \mathbf{W}_b^{(n)}\}$. Note that the variables that are independent of each other can be categorized into one block and optimized together with simple solutions, and the variables that are coupled in the objective function must be divided into different blocks. Therefore, the optimization variables are divided into these three blocks, and we are able to optimize one block of variables among them in each time while keeping the others fixed. This leads to the following three steps in each BCD iteration and as will be shown, in each step closed-form solutions of the associated optimization sub-problems can be obtained.

Step 1: we optimize the variables in \mathcal{B}_1 by fixing \mathcal{B}_2 and \mathcal{B}_3 . We note that problem (24) can be decomposed into five subproblems in this case and each subproblem can be solved independently and in parallel.

$$u_k^{(n)} = \frac{\hat{\mathbf{h}}_k^{(n,n)H} \mathbf{V}^{(n)} \mathbf{a}_k}{\sum_{m \in \mathcal{N}} \|\hat{\mathbf{h}}_k^{(n,m)H} \mathbf{V}^{(m)}\|^2 + \sum_{m \in \mathcal{N}} \sum_{j \in \mathcal{K}} |\hat{c}_{k,j}^{(n,m)}|^2 P_j^{(m)} + \sum_{m \in \mathcal{N}} \sigma_h^2 \|\mathbf{V}^{(m)}\|^2 + \sum_{m \in \mathcal{N}} \sum_{j \in \mathcal{K}} \sigma_c^2 P_j^{(m)} + \sigma_k^2}. \quad (25)$$

(1.1) *The subproblem of $u_k^{(n)}$* : it can be easily seen that this subproblem is an unconstrained quadratic program (QP). By resorting to its first order optimality condition, we can obtain (25), shown at the top of this page.

(1.2) *The subproblem of $\mathbf{U}^{(n)}$* : the variable $\mathbf{U}^{(n)}$ is optimized by solving the following problem:

$$\begin{aligned} \min_{\{\mathbf{U}^{(n)}\}} & \sum_{n \in \mathcal{N}} \{ \text{Tr}(\mathbf{W}_b^{(n)} (\mathbf{U}^{(n)H} \mathbf{R}^{(n)} \mathbf{U}^{(n)} \\ & - \mathbf{U}^{(n)H} \mathbf{G}^{(n,n)} \mathbf{P}^{(n)\frac{1}{2}} - \mathbf{P}^{(n)\frac{1}{2}} \mathbf{G}^{(n,n)} \mathbf{H} \mathbf{U}^{(n)}) \} \\ & + \frac{1}{2\rho} \sum_{n \in \mathcal{N}} (\|\mathbf{U}^{(n)} - \tilde{\mathbf{U}}_{\text{RF}}^{(n)} \mathbf{U}_{\text{BB}}^{(n)} + \rho \lambda_3^{(n)}\|^2), \end{aligned} \quad (26)$$

where

$$\begin{aligned} \mathbf{R}^{(n)} & \triangleq \sum_{m \in \mathcal{N}} \hat{\mathbf{G}}^{(n,m)} \mathbf{P}^{(m)} \hat{\mathbf{G}}^{(n,m)H} + \sum_{m \in \mathcal{N}} \sigma_g^2 \text{Tr}(\mathbf{P}^{(m)}) \mathbf{I} \\ & + \sum_{m \in \mathcal{N} \setminus n} \hat{\mathbf{F}}^{(n,m)} \mathbf{V}^{(m)} \mathbf{V}^{(m)H} \hat{\mathbf{F}}^{(n,m)H} \\ & + \sum_{m \in \mathcal{N} \setminus n} \sigma_f^2 \text{Tr}(\mathbf{V}^{(m)} \mathbf{V}^{(m)H}) \mathbf{I} + \sigma_{\text{BS}}^2 \mathbf{I}. \end{aligned} \quad (27)$$

It can be observed that the above problem is fully decoupled for each n , and it is also an unconstrained QP which can be easily solved. By vectorizing $\mathbf{U}^{(n)}$ and then checking the first order optimality condition, the optimal solution of problem (26) can be obtained as follows:

$$\begin{aligned} \text{vec}(\mathbf{U}^{(n)}) & = \left(\mathbf{W}_b^{(n)T} \otimes \mathbf{R}^{(n)} + \frac{1}{2\rho} \mathbf{I} \right)^{-1} \\ & \left(\text{vec}(\hat{\mathbf{G}}^{(n,m)} \mathbf{P}^{(n)\frac{1}{2}} \mathbf{W}_b^{(n)}) + \frac{1}{2\rho} \text{vec}(\tilde{\mathbf{U}}_{\text{RF}}^{(n)} \mathbf{U}_{\text{BB}}^{(n)} - \rho \lambda_3^{(n)}) \right). \end{aligned} \quad (28)$$

(1.3) *The subproblem of $\mathbf{V}_{\text{BB}}^{(n)}$* : similarly, the optimal $\mathbf{V}_{\text{BB}}^{(n)}$ can be obtained in parallel by

$$\begin{aligned} \mathbf{V}_{\text{BB}}^{(n)} & = \arg \min_{\mathbf{V}_{\text{BB}}^{(n)}} \|\mathbf{V}^{(n)} - \tilde{\mathbf{V}}_{\text{RF}}^{(n)} \mathbf{V}_{\text{BB}}^{(n)} + \rho \lambda_2^{(n)}\|^2 \\ & = (\tilde{\mathbf{V}}_{\text{RF}}^{(n)H} \tilde{\mathbf{V}}_{\text{RF}}^{(n)})^\dagger \left(\tilde{\mathbf{V}}_{\text{RF}}^{(n)H} (\mathbf{V}^{(n)} + \rho \lambda_2^{(n)}) \right). \end{aligned} \quad (29)$$

(1.4) *The subproblem of $\mathbf{V}_{\text{RF}}^{(n)}$* : we have the following optimization problem:

$$\begin{aligned} \min_{\{\mathbf{V}_{\text{RF}}^{(n)}\}} & \sum_{n \in \mathcal{N}} \|\mathbf{V}_{\text{RF}}^{(n)} - \tilde{\mathbf{V}}_{\text{RF}}^{(n)} + \rho \lambda_4^{(n)}\|^2 \\ \text{s.t.} & |\mathbf{V}_{\text{RF}}^{(n)}(i, j)| = 1, \quad \forall i, j, n. \end{aligned} \quad (30)$$

As can be seen, through the introduction of $\tilde{\mathbf{V}}_{\text{RF}}^{(n)}$, problem (30) is made fully separable for different i, j and n , i.e., in this subproblem, all the entries in $\{\mathbf{V}_{\text{RF}}^{(n)}\}$ can be updated simultaneously. Let us express $\tilde{\mathbf{V}}_{\text{RF}}^{(n)}(i, j) - \rho \lambda_4^{(n)}(i, j) = r^{(n)}(i, j) e^{j\phi^{(n)}(i, j)}$, then the optimal solution of problem (30) can be simply obtained as $\mathbf{V}_{\text{RF}}^{(n)}(i, j) = e^{j\phi^{(n)}(i, j)}$. In other

words, the phases of $\mathbf{V}_{\text{RF}}^{(n)}$ should be aligned with those of $\tilde{\mathbf{V}}_{\text{RF}}^{(n)} - \rho \lambda_4^{(n)}$ in element-wise.

(1.5) *The subproblem of $\tilde{\mathbf{V}}^{(n)}$* : the variable $\tilde{\mathbf{V}}^{(n)}$ can be updated by solving

$$\begin{aligned} \min_{\tilde{\mathbf{V}}^{(n)}} & \|\mathbf{V}^{(n)} - \tilde{\mathbf{V}}^{(n)} + \rho \lambda_1^{(n)}\|^2 \\ \text{s.t.} & \|\tilde{\mathbf{V}}^{(n)}\|^2 \leq P_B. \end{aligned} \quad (31)$$

Note that solving (31) is to find a point on a sphere centered at the origin that minimizes the distance to the point $\mathbf{V}^{(n)} + \rho \lambda_1^{(n)}$. Thus, its optimal solution can be obtained based on projection, which is the following closed-form solution:

$$\tilde{\mathbf{V}}^{(n)} = \mathbb{P}_{\Omega_1} \{\mathbf{V}^{(n)} + \rho \lambda_1^{(n)}\}, \quad (32)$$

where $\Omega_1 \triangleq \{\tilde{\mathbf{V}}^{(n)} \mid \|\tilde{\mathbf{V}}^{(n)}\|^2 \leq P_B\}$.

Step 2: in this step, we consider the optimization of \mathcal{B}_2 when all the other variables are fixed. Similar to Step 1, in this case, problem (24) can be divided into four independent subproblems and each subproblem can be efficiently solved as will be shown below.

(2.1) *The subproblem of $\mathbf{V}^{(n)}$* : the variable $\mathbf{V}^{(n)}$ can be obtained by solving the following unconstrained QP:

$$\begin{aligned} \min_{\mathbf{V}^{(n)}} & \sum_{m \in \mathcal{N}} \sum_{k \in \mathcal{K}} w_k^{(m)} |u_k^{(m)}|^2 \left(\|\hat{\mathbf{h}}_k^{(m,n)H} \mathbf{V}^{(n)}\|^2 + \sigma_h^2 \|\mathbf{V}^{(n)}\|^2 \right) \\ & - \sum_{k \in \mathcal{K}} 2w_k^{(n)} \Re(u_k^{(n)*} \hat{\mathbf{h}}_k^{(n,n)H} \mathbf{V}^{(n)} \mathbf{a}_k) \\ & + \sum_{m \in \mathcal{N} \setminus n} \text{Tr}(\mathbf{W}_b^{(m)} \mathbf{U}^{(m)H} (\hat{\mathbf{F}}^{(m,n)} \mathbf{V}^{(n)} \mathbf{V}^{(n)H} \hat{\mathbf{F}}^{(m,n)H} \\ & + \sigma_f^2 \text{Tr}(\mathbf{V}^{(n)} \mathbf{V}^{(n)H}) \mathbf{I}) \mathbf{U}^{(m)}) + \frac{1}{2\rho} (\|\mathbf{V}^{(n)} - \tilde{\mathbf{V}}^{(n)} + \rho \lambda_1^{(n)}\|^2 \\ & + \|\mathbf{V}^{(n)} - \tilde{\mathbf{V}}_{\text{RF}}^{(n)} \mathbf{V}_{\text{BB}}^{(n)} + \rho \lambda_2^{(n)}\|^2), \end{aligned} \quad (33)$$

whose optimal solution is given by (by resorting to the first-order optimality condition)

$$\begin{aligned} \mathbf{V}^{(n)} & = (\mathbf{Z}^{(n)})^{-1} \left(w_k^{(n)} u_k^{(n)} \hat{\mathbf{h}}_k^{(n,n)} \mathbf{a}_k^T \right. \\ & \left. + \frac{1}{2\rho} (\tilde{\mathbf{V}}^{(n)} - \rho \lambda_1^{(n)} + \tilde{\mathbf{V}}_{\text{RF}}^{(n)} \mathbf{V}_{\text{BB}}^{(n)} - \rho \lambda_2^{(n)}) \right), \end{aligned} \quad (34)$$

and

$$\begin{aligned} \mathbf{Z}^{(n)} & = \sum_{m \in \mathcal{N}} \sum_{k \in \mathcal{K}} w_k^{(m)} |u_k^{(m)}|^2 (\hat{\mathbf{h}}_k^{(m,n)} \hat{\mathbf{h}}_k^{(m,n)H} + \sigma_h^2 \mathbf{I}) \\ & + \sum_{m \in \mathcal{N} \setminus n} \hat{\mathbf{F}}^{(m,n)H} \mathbf{U}^{(m)} \mathbf{W}_b^{(m)} \mathbf{U}^{(m)H} \hat{\mathbf{F}}^{(m,n)} \\ & + \sum_{m \in \mathcal{N} \setminus n} \sigma_f^2 \text{Tr}(\mathbf{U}^{(m)} \mathbf{W}_b^{(m)} \mathbf{U}^{(m)H}) \mathbf{I} + \frac{1}{\rho} \mathbf{I}. \end{aligned} \quad (35)$$

(2.2) *The subproblem of $\mathbf{P}^{(n)}$* : extracting the terms in problem (24) involving $\mathbf{P}^{(n)}$ results to problem (36), shown at the top of the next page. By making the following variable

$$\begin{aligned}
& \min_{\{\mathbf{P}^{(n)}\}} \sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} \left(w_k^{(n)} (|u_k^{(n)}|^2 \left(\sum_{m \in \mathcal{N}} \sum_{j \in \mathcal{K}} |\hat{c}_{k,j}^{(n,m)}|^2 P_j^{(m)} + \sum_{m \in \mathcal{N}} \sum_{j \in \mathcal{K}} \sigma_c^2 P_j^{(m)} \right)) \right. \\
& \quad + \sum_{n \in \mathcal{N}} \left(\text{Tr} \left(\mathbf{W}_b^{(n)} \left(\mathbf{U}^{(n)H} \left(\sum_{m \in \mathcal{N}} \hat{\mathbf{G}}^{(n,m)} \mathbf{P}^{(m)} \hat{\mathbf{G}}^{(n,m)H} + \sum_{m \in \mathcal{N}} \sigma_g^2 \text{Tr}(\mathbf{P}^{(m)}) \mathbf{I} \right) \mathbf{U}^{(n)} \right. \right. \\
& \quad \left. \left. - \mathbf{U}^{(n)H} \mathbf{G}^{(n,n)} \mathbf{P}^{(n)\frac{1}{2}} - \mathbf{P}^{(n)\frac{1}{2}} \mathbf{G}^{(n,n)H} \mathbf{U}^{(n)} \right) \right) \left. \right) \\
& \quad \text{s.t. } P_j^{(n)} \leq P_u, \forall j, n.
\end{aligned} \tag{36}$$

substitution, i.e., $q_j^{(n)} \triangleq \sqrt{P_j^{(n)}}$, it leads to an equivalent problem given by:

$$\begin{aligned}
& \min_{q_j^{(n)}} q_{d,j}^{(n)} q_j^{(n)2} + [\mathbf{W}_b^{(n)} \mathbf{U}^{(n)H} \hat{\mathbf{G}}^{(n,n)} \\
& \quad + \hat{\mathbf{G}}^{(n,n)H} \mathbf{U}^{(n)} \mathbf{W}_b^{(n)}]_{jj} q_j^{(n)} \\
& \quad \text{s.t. } q_j^{(n)2} \leq P_u,
\end{aligned} \tag{37}$$

where $q_{d,j}^{(n)} \triangleq \sum_{m \in \mathcal{N}} \sum_{k \in \mathcal{K}} w_k^{(m)} |u_k^{(m)}|^2 (|\hat{c}_{k,j}^{(m,n)}|^2 + \sigma_c^2) + \sum_{m \in \mathcal{N}} [\hat{\mathbf{G}}^{(m,n)H} \mathbf{U}^{(m)} \mathbf{W}_b^{(m)} \mathbf{U}^{(m)H} \hat{\mathbf{G}}^{(m,n)}]_{jj} + \sum_{m \in \mathcal{N}} \text{Tr}(\mathbf{U}^{(m)} \mathbf{W}_b^{(m)} \mathbf{U}^{(m)H}) \sigma_g^2$ and this problem is naturally decoupled among different j and n . Problem (37) is a quadratic constrained quadratic programming (QCQP) problem with only one constraint, whose optimal solution can be obtained in closed-form. The details are relegated to Appendix B.

(2.3) *The subproblem of $\mathbf{U}_{BB}^{(n)}$* : similar to subproblem (1.3), we have the following problem:

$$\min_{\mathbf{U}_{BB}^{(n)}} \|\mathbf{U}^{(n)} - \tilde{\mathbf{U}}_{\text{RF}}^{(n)} \mathbf{U}_{\text{BB}}^{(n)} + \rho \boldsymbol{\lambda}_3^{(n)}\|^2, \tag{38}$$

and the optimal solution can be obtained accordingly, which can be expressed as

$$\mathbf{U}_{\text{BB}}^{(n)} = (\tilde{\mathbf{U}}_{\text{RF}}^{(n)H} \tilde{\mathbf{U}}_{\text{RF}}^{(n)})^\dagger \left(\tilde{\mathbf{U}}_{\text{RF}}^{(n)H} (\mathbf{U}^{(n)} + \rho \boldsymbol{\lambda}_3^{(n)}) \right). \tag{39}$$

(2.4) *The $\mathbf{U}_{\text{RF}}^{(n)}$ subproblem*: this subproblem can be similarly addressed by following the derivation in (1.4), therefore the details are omitted here for brevity.

Step 3: we introduce the optimization of \mathcal{B}_3 by fixing the other variables. Accordingly, we have the following two subproblems:

(3.1) *The subproblems of $w_k^{(n)}$ and $\mathbf{W}_b^{(n)}$* : we can see that the objective function in (23) is convex w.r.t. $w_k^{(n)}$ and $\mathbf{W}_b^{(n)}$ by keeping the other variables fixed. As a result, by checking the corresponding first order optimality conditions, we obtain

$$w_k^{(n)} = \frac{1}{\tilde{e}_k^{(n)}}, \quad \mathbf{W}_b^{(n)} = (\tilde{\mathbf{E}}_b^{(n)})^{-1}, \tag{40}$$

where $\tilde{e}_k^{(n)}$ and $\tilde{\mathbf{E}}_b^{(n)}$ are obtained by replacing $\mathbf{V}_{\text{RF}}^{(n)} \mathbf{V}_{\text{BB}}^{(n)}$ and $\mathbf{U}_{\text{RF}}^{(n)} \mathbf{U}_{\text{BB}}^{(n)}$ with $\mathbf{V}^{(n)}$ and $\mathbf{U}^{(n)}$ in (20) and (21), respectively.

(3.2) *The subproblems of $\tilde{\mathbf{U}}_{\text{RF}}^{(n)}$ and $\tilde{\mathbf{V}}_{\text{RF}}^{(n)}$* : since the optimization subproblems of $\tilde{\mathbf{U}}_{\text{RF}}^{(n)}$ and $\tilde{\mathbf{V}}_{\text{RF}}^{(n)}$ have similar structures, we focus on the update of $\tilde{\mathbf{U}}_{\text{RF}}^{(n)}$ and the solution of $\tilde{\mathbf{V}}_{\text{RF}}^{(n)}$

can be obtained similarly. The subproblem of $\tilde{\mathbf{U}}_{\text{RF}}^{(n)}$ is given by

$$\begin{aligned}
& \min_{\tilde{\mathbf{U}}_{\text{RF}}^{(n)}} \|\mathbf{U}^{(n)} - \tilde{\mathbf{U}}_{\text{RF}}^{(n)} \mathbf{U}_{\text{BB}}^{(n)} + \rho \boldsymbol{\lambda}_3^{(n)}\|^2 \\
& \quad + \|\mathbf{U}_{\text{RF}}^{(n)} - \tilde{\mathbf{U}}_{\text{RF}}^{(n)} + \rho \boldsymbol{\lambda}_5^{(n)}\|^2,
\end{aligned} \tag{41}$$

whose optimal solution is $\tilde{\mathbf{U}}_{\text{RF}}^{(n)} = (\mathbf{U}^{(n)} \mathbf{U}_{\text{BB}}^{(n)H} + \mathbf{U}_{\text{RF}}^{(n)} - \rho \boldsymbol{\lambda}_3^{(n)} \mathbf{U}_{\text{BB}}^{(n)H} - \rho \boldsymbol{\lambda}_5^{(n)}) (\mathbf{U}_{\text{BB}}^{(n)} \mathbf{U}_{\text{BB}}^{(n)H} + \mathbf{I})^{-1}$.

To sum up, the above three updating steps are successively performed in each iteration of the BCD method, according to the block structure of the optimization variables. Therefore, the inner loop of the proposed PDD-based algorithm (i.e., the proposed BCD method) is summarized in **Algorithm 2**. In the outer loop, after running **Algorithm 2**, we update the dual variables $\{\boldsymbol{\lambda}_i\}, i \in \{1, \dots, 5\}$ or the penalty parameter ρ according to the following constraint violation (CV) condition: $\max_{n \in \mathcal{N}} \{\max\{\|\mathbf{V}^{(n)} - \tilde{\mathbf{V}}_{\text{RF}}^{(n)} \mathbf{V}_{\text{BB}}^{(n)}\|_\infty, \|\mathbf{U}^{(n)} - \tilde{\mathbf{U}}_{\text{RF}}^{(n)} \mathbf{U}_{\text{BB}}^{(n)}\|_\infty, \|\tilde{\mathbf{V}}^{(n)} - \mathbf{V}^{(n)}\|_\infty, \|\mathbf{V}_{\text{RF}}^{(n)} - \tilde{\mathbf{V}}_{\text{RF}}^{(n)}\|_\infty, \|\mathbf{U}_{\text{RF}}^{(n)} - \tilde{\mathbf{U}}_{\text{RF}}^{(n)}\|_\infty\}\} \leq \eta^r$ (see Step 4 in **Algorithm 1**). Specifically, the dual variables are updated according to

$$\begin{aligned}
\boldsymbol{\lambda}_{1,m+1}^{(n)} &= \boldsymbol{\lambda}_{1,m}^{(n)} + \frac{1}{\rho} (\mathbf{V}^{(n)} - \tilde{\mathbf{V}}^{(n)}), \quad \forall n, \\
\boldsymbol{\lambda}_{2,m+1}^{(n)} &= \boldsymbol{\lambda}_{2,m}^{(n)} + \frac{1}{\rho} (\mathbf{V}^{(n)} - \tilde{\mathbf{V}}_{\text{RF}}^{(n)} \mathbf{V}_{\text{BB}}^{(n)}), \quad \forall n, \\
\boldsymbol{\lambda}_{3,m+1}^{(n)} &= \boldsymbol{\lambda}_{3,m}^{(n)} + \frac{1}{\rho} (\mathbf{U}^{(n)} - \tilde{\mathbf{U}}_{\text{RF}}^{(n)} \mathbf{U}_{\text{BB}}^{(n)}), \quad \forall n, \\
\boldsymbol{\lambda}_{4,m+1}^{(n)} &= \boldsymbol{\lambda}_{4,m}^{(n)} + \frac{1}{\rho} (\mathbf{V}_{\text{RF}}^{(n)} - \tilde{\mathbf{V}}_{\text{RF}}^{(n)}), \quad \forall n, \\
\boldsymbol{\lambda}_{5,m+1}^{(n)} &= \boldsymbol{\lambda}_{5,m}^{(n)} + \frac{1}{\rho} (\mathbf{U}_{\text{RF}}^{(n)} - \tilde{\mathbf{U}}_{\text{RF}}^{(n)}), \quad \forall n,
\end{aligned} \tag{42}$$

if the CV condition is satisfied, otherwise, we decrease the penalty parameter by Step 7 in **Algorithm 1**, where $0 < c < 1$ is a parameter which controls the decreasing speed of ρ^r . Note that when ρ^r is sufficiently small, it will force the penalized constraints to be satisfied with equality after solving problem (24). In other words, as the iteration of the proposed PDD-based algorithm increases, a feasible solution to problem (23) can be obtained. Moreover, we have the following proposition for problem (23):

Proposition III.2. *Mangasarian-Fromovitz constraint qualification (MFCQ) condition is satisfied for problem (23) at any feasible point $\Delta_{\varphi^{(n)}}$.*

Proof. The detailed proof is relegated to Appendix C. \square

Therefore, using Proposition III.2 and following Theorem 3.1 in [47], it can be proved that the proposed PDD-based algorithm is able to converge to a KKT solution of problem (23). Furthermore, due to Proposition III.1, we can conclude

that the proposed algorithm also converges to the set of KKT solutions of problem (15).

Algorithm 2 Proposed BCD method for problem (24)

- 1: Define the maximum iteration number L_{\max} and the tolerance of accuracy ϵ . Initialize with a feasible point and set the iteration number $l = 0$.
 - 2: **repeat**
 - 3: Update \mathcal{B}_1 , \mathcal{B}_2 and \mathcal{B}_3 successively in **Steps 1-3**.
 - 4: $l = l + 1$.
 - 5: **until** The fractional difference of the objective value is less than ϵ or the maximum number of iterations is reached, i.e., $l \geq L_{\max}$.
-

D. Complexity Analysis

In this subsection, we aim to analyze the computational complexity of the proposed PDD-based algorithm. Specifically, the computational complexity of the subproblems in Steps 1-3 of the inner BCD method is listed in Table I, where solving subproblem (1.2) is the most costly step in terms of complexity order which requires $\mathcal{O}(N(N_r J)^3)$. On this basis, we are able to analyze the asymptotic complexity of the proposed algorithm when N_t , N_r , K , J , N_{RF} and N are close to infinity. Under this condition, one can verify that the complexity of the proposed algorithm is on the order of $\mathcal{O}(I_1 I_2 N N_r^3 J^3)$, where I_1 and I_2 denote the total number of outer and inner iterations needed by the proposed algorithm.

TABLE I: Complexity analysis

Subproblem		Complexity
Step 1	$u_k^{(n)}$	$\mathcal{O}(N^2 N_t K)$
	$\mathbf{U}^{(n)}$	$\mathcal{O}(N(N_r J)^3)$
	$\mathbf{V}_{\text{BB}}^{(n)}$	$\mathcal{O}(N(N_{\text{RF}}^3 + N_{\text{RF}}^2 N_t))$
	$\mathbf{V}_{\text{RF}}^{(n)}$	$\mathcal{O}(N N_t N_{\text{RF}})$
	$\tilde{\mathbf{V}}^{(n)}$	$\mathcal{O}(N N_t K)$
Step 2	$\mathbf{V}^{(n)}$	$\mathcal{O}(N N_t^3)$
	$\mathbf{P}^{(n)}$	$\mathcal{O}(N N_r^2 J)$
	$\mathbf{U}_{\text{BB}}^{(n)}$	$\mathcal{O}(N(N_{\text{RF}}^2 N_r + N_{\text{RF}}^3))$
	$\mathbf{U}_{\text{RF}}^{(n)}$	$\mathcal{O}(N N_r N_{\text{RF}})$
Step 3	$w_k^{(n)}$	$\mathcal{O}(N^2 N_t^2 K)$
	$\mathbf{W}_b^{(n)}$	$\mathcal{O}(N(N_r^2 K + J^3))$
	$\tilde{\mathbf{U}}_{\text{RF}}^{(n)}$	$\mathcal{O}(N N_{\text{RF}}^3)$
	$\tilde{\mathbf{V}}_{\text{RF}}^{(n)}$	$\mathcal{O}(N N_{\text{RF}}^3)$

IV. EXTENSION TO THE SUBARRAY STRUCTURE

In the previous section, we investigate the robust joint transceiver algorithm based on a fully-connected hybrid AD processing structure, i.e., each RF chain is connected to all the antennas. Such hybrid beamforming scheme exhibits lower hardware complexity compared to that of a fully digital structure. However, its complexity can be further reduced by connecting each RF chain to a subset of antennas [1]. Thus, this subarray structure offers a potential tradeoff between implementation cost and performance for large-scale antenna arrays and therefore serves as a promising solution for future mmWave cellular communications [1], [51]. In this section,

we extend the proposed PDD-based algorithm to the subarray structure, and show that the proposed design methodology exhibits high flexibility for different hybrid processing structures.

In the subarray structure, the FD BS in each cell contains N_{RF} RF chains for transmit and receive processing, respectively, and each transmit (receive) RF chain is connected to $L_V \triangleq \frac{N_t}{N_{\text{RF}}}$ ($L_U \triangleq \frac{N_r}{N_{\text{RF}}}$) antennas.² As a result, the transmit and receive analog beamforming matrix $\mathbf{V}_{\text{RF}}^{(n)} \in \mathbb{C}^{N_{\text{RF}} L_V \times N_{\text{RF}}}$ and $\mathbf{U}_{\text{RF}}^{(n)} \in \mathbb{C}^{N_{\text{RF}} L_U \times N_{\text{RF}}}$ shows a block diagonal structure, which can be expressed as $\mathbf{V}_{\text{RF}}^{(n)} = \text{blkdiag}\{\mathbf{v}_{\text{RF},1}^{(n)}, \dots, \mathbf{v}_{\text{RF},N_{\text{RF}}}^{(n)}\}$ and $\mathbf{U}_{\text{RF}}^{(n)} = \text{blkdiag}\{\mathbf{u}_{\text{RF},1}^{(n)}, \dots, \mathbf{u}_{\text{RF},N_{\text{RF}}}^{(n)}\}$, where $\mathbf{v}_{\text{RF},i}^{(n)} \in \mathbb{C}^{L_V \times 1}$ and $\mathbf{u}_{\text{RF},i}^{(n)} \in \mathbb{C}^{L_U \times 1}$ are the analog beamforming vectors, the entries of which follow

$$|\mathbf{v}_{\text{RF},i}^{(n)}(l_v)| = 1, \forall i \in \{1, \dots, N_{\text{RF}}\}, l_v \in \{1, \dots, L_V\} \quad (43)$$

and

$$|\mathbf{u}_{\text{RF},i}^{(n)}(l_u)| = 1, \forall i \in \{1, \dots, N_{\text{RF}}\}, l_u \in \{1, \dots, L_U\}, \quad (44)$$

respectively. Consequently, the corresponding AL problem can be formulated as

$$\begin{aligned} & \min_{\{\varphi^{(n)}, u_k^{(n)}, w_k^{(n)}, \mathbf{W}_b^{(n)}\}} \sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} \tilde{h}^{(n,k)}(u_k^{(n)}, w_k^{(n)}, \varphi^{(n)}) \\ & + \sum_{n \in \mathcal{N}} \tilde{g}^{(n)}(\mathbf{W}_b^{(n)}, \varphi^{(n)}) \\ & + \frac{1}{2\rho} \sum_{n \in \mathcal{N}} \left(+ \|\mathbf{V}^{(n)} - \tilde{\mathbf{V}}_{\text{RF}}^{(n)} \mathbf{V}_{\text{BB}}^{(n)} + \rho \lambda_2^{(n)}\|^2 \right. \\ & + \|\mathbf{U}^{(n)} - \tilde{\mathbf{U}}_{\text{RF}}^{(n)} \mathbf{U}_{\text{BB}}^{(n)} + \rho \lambda_3^{(n)}\|^2 + \|\mathbf{V}^{(n)} - \tilde{\mathbf{V}}^{(n)} + \rho \lambda_1^{(n)}\|^2 \\ & \left. + \|\mathbf{V}_{\text{RF}}^{(n)} - \tilde{\mathbf{V}}_{\text{RF}}^{(n)} + \rho \lambda_4^{(n)}\|^2 + \|\mathbf{U}_{\text{RF}}^{(n)} - \tilde{\mathbf{U}}_{\text{RF}}^{(n)} + \rho \lambda_5^{(n)}\|^2 \right) \\ & \text{s.t. (23b), (23c), (43) and (44).} \end{aligned} \quad (45)$$

Note that when develop a BCD method for solving problem (45) in the inner loop of the PDD-based algorithm, the updating rules of most of the steps are similar to the fully-connected case. The main difference lies in solving the subproblems of $\mathbf{U}_{\text{RF}}^{(n)}$ and $\mathbf{V}_{\text{RF}}^{(n)}$. For illustration, we only consider the updating of $\mathbf{U}_{\text{RF}}^{(n)}$ in the following, and the updating of $\mathbf{V}_{\text{RF}}^{(n)}$ can be addressed similarly. The subproblem of $\mathbf{U}_{\text{RF}}^{(n)}$ can be expressed as

$$\begin{aligned} & \min_{\mathbf{U}_{\text{RF}}^{(n)}} \sum_{i=1}^{N_{\text{RF}}} \|\mathbf{u}_{\text{RF},i}^{(n)} - \tilde{\mathbf{u}}_{\text{RF},i}^{(n)} + \rho \lambda_{5,i}^{(n)}\|^2 \\ & \text{s.t. } |\mathbf{u}_{\text{RF},i}^{(n)}(l_u)| = 1, \forall i, l_u, \end{aligned} \quad (46)$$

where $\tilde{\mathbf{u}}_{\text{RF},i}^{(n)}$ and $\lambda_{5,i}^{(n)}$ are $L_U \times 1$ vectors that are taken from the $((i-1)L_U + 1)$ -th to the (iL_U) -th rows and the i -th column of the matrices $\tilde{\mathbf{U}}_{\text{RF}}^{(n)}$ and $\lambda_5^{(n)}$, respectively. By using a similar technique as that in subproblem (1.4) of Section III, we can observe that problem (46) can be globally solved and the details are omitted here for conciseness. In this subproblem, only $N_{\text{RF}} L_U$ elements in $\mathbf{U}_{\text{RF}}^{(n)}$ need to be optimized, and the remaining elements can be directly set to zeros.

Furthermore, similar to the complexity analysis in Section III-D, the asymptotic complexity of the proposed PDD-based algorithm in the subarray structure can be written

²For simplicity, we assume that N_t and N_r are divisible by N_{RF} .

as $\mathcal{O}\left(I_1 I_2 N (L_U N_{RF} J)^3\right)$. Note that although the proposed algorithm in the fully-connected and subarray structures exhibit the same complexity order, utilizing the subarray structure would be much simpler in practice since the variables $\{\mathbf{V}_{RF}^{(n)}, \mathbf{U}_{RF}^{(n)}, \tilde{\mathbf{V}}_{RF}^{(n)}, \tilde{\mathbf{U}}_{RF}^{(n)}\}$ admit a sparse structure.

V. SIMULATION RESULTS

In this section, we present numerical results to evaluate the performance of the proposed PDD-based robust joint hybrid transceiver algorithm. We employ a geometric channel model of $L = 15$ paths with uniform linear array antenna configurations, which is widely used in mmWave [9], [12], [16]. To be specific, the channel matrices between the BSs in different cells are expressed as

$$\mathbf{F}^{(n,m)} = \sqrt{\frac{N_t N_r}{L}} \sum_{l=1}^L \alpha_l^{(n,m)} \mathbf{a}_r^{(n,m)}(\phi_l^{(n,m)}) \mathbf{a}_t^{(n,m)}(\varphi_l^{(n,m)})^H, \quad \forall n, m, \quad (47)$$

where $\alpha_l^{(n,m)} \sim \mathcal{CN}(0, 1)$ is the complex gain of the l -th path, $\phi_l^{(n,m)} \in [0, 2\pi)$ and $\varphi_l^{(n,m)} \in [0, 2\pi)$ are uniformly and randomly generated. Further, $\mathbf{a}_r^{(n,m)}(\phi_l^{(n,m)})$ and $\mathbf{a}_t^{(n,m)}(\varphi_l^{(n,m)})$ are the normalized array response vectors at the transmitter and receiver, respectively, which are given by

$$\mathbf{a}(\theta) = \frac{1}{N} [1, e^{j\pi \sin(\theta)}, \dots, e^{j(N-1)\pi \sin(\theta)}]^T. \quad (48)$$

The channel matrices $\mathbf{H}^{(n,m)}$ and $\mathbf{G}^{(n,m)}$ are generated in a similar way, and the channel coefficients $\{c_{k,j}^{(n,m)}\}$ between the users are assumed to follow Rayleigh fading.

We assume that there are $N = 2$ cells in the system, each cell contains $K = 4$ DL users and $J = 3$ UL users, each BS is equipped with $N_t = N_r = 48$ transmit/receive antennas and $N_{RF} = 12$ RF chains unless otherwise specified. For simplicity, the noise parameters are set as $\sigma_k^2 = \sigma_{BS}^2 = 0.01$ and we let $\sigma_e^2 = \sigma_g^2 = \sigma_h^2 = \sigma_f^2 = \sigma_c^2$. For the proposed PDD-based algorithm, the initial penalty parameter ρ^0 is set to 5 and the corresponding control parameter c is set to 0.95. Furthermore, we set $\eta^r = \sqrt{\zeta^r}$, $\zeta^0 = 0.01$ and $\zeta^{r+1} = 0.8\zeta^r$. Moreover, for practical implementation, we set the maximum inner iteration number to 50 and the maximum outer iteration number to 300. Note that when we already have enough constraint violation accuracy, the convergence of the inner loop algorithm might become slow due to a small penalty parameter. Thus, by setting a maximum iteration number, we can save computational resources without sacrificing any performance based on our experience. For a given SNR value, the power budgets of the UL users and the BS are set to $P_u = 10^{\text{SNR}/10} \sigma_k^2$ and $P_B = 10^{\text{SNR}/10} K \sigma_k^2$. For comparison, we provide the performance of the fully digital transceiver scheme as a benchmark. Moreover, a separate design scheme is simulated where the analog beamforming matrices are randomly generated (or by the singular vectors of the channel matrices through singular value decomposition (SVD)), while the digital beamforming matrices are optimized using the PDD-based algorithm. The considered separate design algorithm with the aforementioned two different analog beamforming

design strategies are referred to as *Random FD* and *SVD FD* in the following, respectively. The computational complexity orders of the fully digital and separate design schemes are on the order of $\mathcal{O}(I_1 I_2 N N_r^3 J^3)$, since they are also dominated by the procedure of updating $\mathbf{U}^{(n)}$.

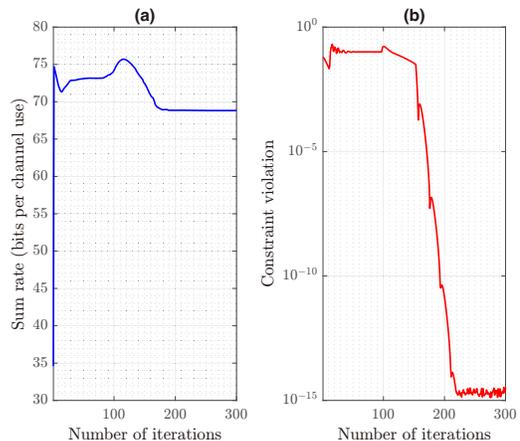


Fig. 2: (a) Sum rate and (b) CV versus the iteration number for the proposed PDD-based algorithm with SNR = 10 dB and $\sigma_e^2 = 0.04$.

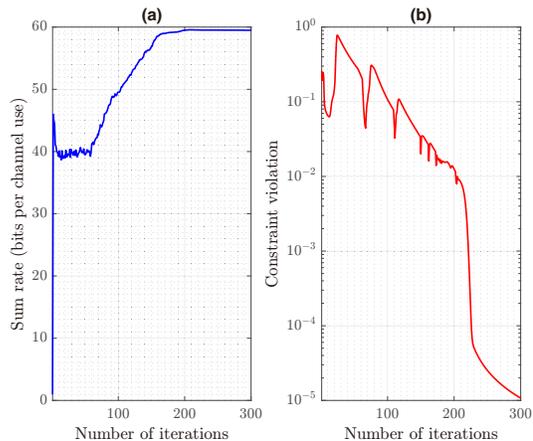


Fig. 3: (a) Sum rate and (b) CV versus the iteration number for the PDD-based algorithm in the subarray structure with SNR = 10 dB and $\sigma_e^2 = 0.04$.

In Fig. 2 and Fig. 3, we illustrate the convergence behaviors of the proposed PDD-based robust joint hybrid transceiver algorithm under the fully-connected and subarray structures. It can be observed from the plots that the proposed PDD-based algorithm can converge well within a few hundreds of iterations for both structures. Since it is not a *monotonic* convergent iterative algorithm, fluctuations of the objective value in the initial few iterations are expected. As the number of iterations increases, the objective value tends to be convergent, as can be seen from Fig. 2 (a) and Fig. 3 (a). This phenomenon can be interpreted as follows: since the initial penalty parameter is relatively large, thereby the solutions obtained are not feasible solutions yet and this results in the oscillatory behavior. Furthermore, from Fig. 2 (b) and Fig. 3

(b), we can see that the CV tends to be retained at relatively high values at first, when the penalty parameter gradually decreases (i.e., the penalty $\frac{1}{2\rho}$ increases) and reaches a proper value, the CV would drop to a practically acceptable level rapidly, e.g., below 10^{-5} .

Fig. 4 shows the average sum rate performance versus different SNR levels with various numbers of RF chains, where we set the CSI error variance to $\sigma_e^2 = 0.04$. We can observe from this figure that the performance of the proposed PDD-based algorithm with FD BSs is much better than that of its HD counterpart with HD BSs, and the sum rate gain becomes larger for higher SNR levels. For both FD and HD schemes, their performance improves with the increasing of RF chains and SNR levels. However, there is generally a performance limit with a given number of RF chains, e.g., the sum rate performance of the FD scheme is near saturation when the SNR level is above 20 dB with $N_{\text{RF}} = 4$. Moreover, the FD scheme with $N_{\text{RF}} = 10$ is able to achieve a very similar performance compared with the benchmark.

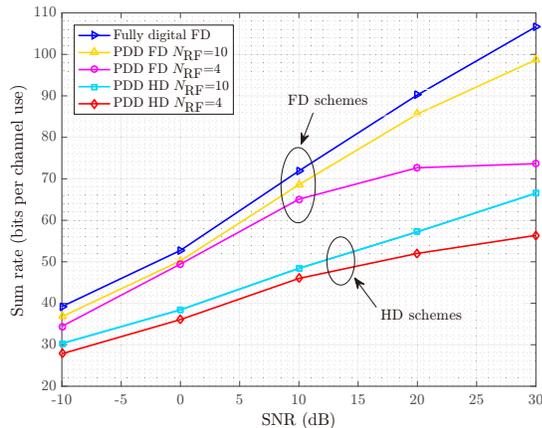


Fig. 4: Sum rate performance comparison between the FD and HD schemes with $\sigma_e^2 = 0.04$.

Next, we compare the performance of the robust and non-robust designs, in terms of the average achievable sum rate. The results are presented in Fig. 5, where the performance is compared with different CSI error variances, i.e., σ_e^2 . We can observe that the performance of the proposed robust PDD-based algorithm and its nonrobust counterpart become closer to that of the perfect CSI case as σ_e^2 decreases. However, the robust design algorithm significantly outperforms its nonrobust counterpart. Specifically, around 9 bits per channel use performance gain can be achieved at $\sigma_e^2 = 0.04$. Furthermore, this performance gain gets larger with the increasing of σ_e^2 , which shows the superiority of the proposed PDD-based algorithm.

In Fig. 6 and Fig. 7, we investigate the sum rate performance of the proposed algorithm versus different numbers of RF chains and cells. One can observe from Fig. 6 that when $N_{\text{RF}} \geq 20$, the performance of the proposed algorithm would be near optimal, which significantly outperforms the separate design algorithm since the latter generally requires $N_{\text{RF}} \approx (N_t = N_r)$ in order to obtain near optimal performance. Moreover, from Fig. 7 we can observe that the average sum rate performance improves with the increasing of the number

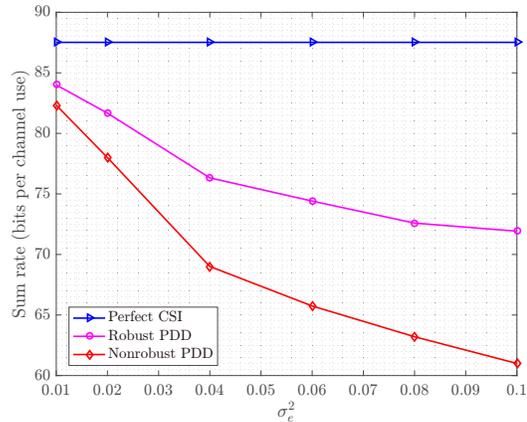


Fig. 5: Sum rate performance comparison between the robust and nonrobust designs with $N_{\text{RF}} = 10$.

of cells. The performance gaps between PDD and the separate design algorithms enlarge with more cells, however, the sum rate growth rate gradually decreases due to the fact that more cells potentially bring more interference.

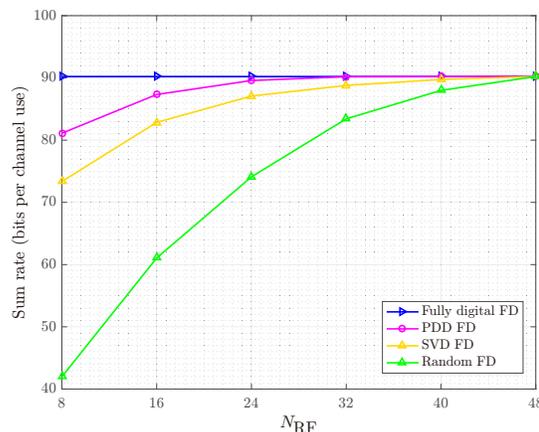


Fig. 6: Sum rate performance versus the number of RF chains for the proposed PDD-based algorithm and the separate design algorithm with $\text{SNR} = 20$ dB and $\sigma_e^2 = 0.04$.

In Fig. 8, we demonstrate the sum rate performance achieved by the proposed algorithm with both fully-connected and subarray structures in the finite resolution phase-shifter case. It can be observed that, the sum rate performance of the proposed algorithm improves when more bits are used for phase quantization, and this improvement shrinks as the quantization level increases (i.e., with the increasing of quantization bits B). Particularly, with the fully-connected structure, we can almost achieve the performance of $B = \infty$ using eight bits quantization. Four bits quantization is practically enough for achieving most of the performance gain in this considered example. While for the case of subarray structure, the number of quantization bits can be reduced to four in order to achieve negligible performance loss.

Finally, in Fig. 9, we investigate the rate performance of the UL and DL achieved by the proposed algorithm versus

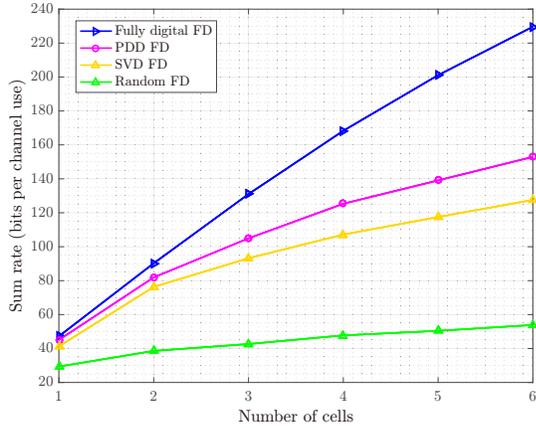


Fig. 7: Sum rate performance versus the number of cells for the proposed PDD-based algorithm and the separate design algorithm with $\text{SNR} = 20$ dB, $N_{\text{RF}} = 10$ and $\sigma_e^2 = 0.04$.

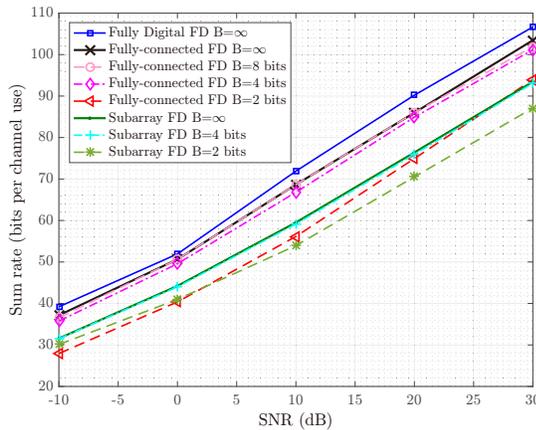


Fig. 8: Sum rate performance comparison between the fully-connected and subarray structures with $\text{SNR} = 20$ dB, $N_{\text{RF}} = 12$ and $\sigma_e^2 = 0.04$.

the UL power budget P_u . As can be seen, when P_u is small (e.g. $P_u = 0.05 \times 10^{\text{SNR}/10} \sigma_k^2$), the average sum rates of the UL and DL are similar, however, when P_u increases (e.g. $P_u = 10^{\text{SNR}/10} \sigma_k^2$), the average sum rate of the UL becomes noticeably larger than that of the DL. This is mainly due to the fact that for the DL, the UL-to-DL interference in (1) is positively related to P_u and cannot be effectively canceled since the DL users are only equipped with a single antenna. However, for the UL, the impact of the DL-to-UL interference in (3) can be compensated by properly designing $\{\mathbf{V}_{\text{RF}}^{(n)}, \mathbf{V}_{\text{BB}}^{(n)}, \mathbf{U}_{\text{RF}}^{(n)}, \mathbf{U}_{\text{BB}}^{(n)}\}$ and the useful signal power gets larger with the increasing of P_u .

VI. CONCLUSION

In this paper, we considered the robust joint hybrid transceiver design problem for a FD mmWave multicell system with imperfect CSI. To mitigate the incurred complex interference, we formulated a worst-case sum rate maximization problem by taking stochastic CSI errors into consideration. To address this difficult problem, we proposed to first transform

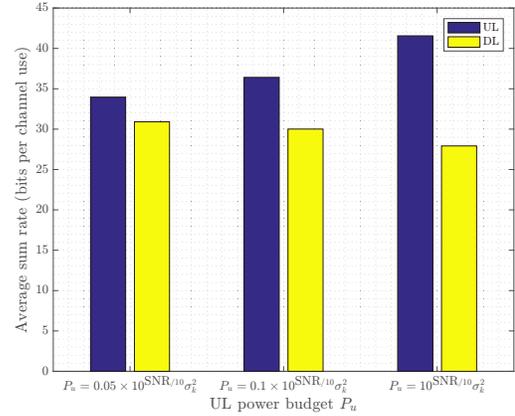


Fig. 9: Sum rate performance comparison between the UL and DL with $\text{SNR} = 15$ dB, $N_{\text{RF}} = 4$ and $\sigma_e^2 = 0.04$.

it into an equivalent but more tractable form by properly introducing some auxiliary variables and employing the WMMSE method. Then, a PDD-based algorithm was developed to iteratively solve this equivalent problem. Especially, we showed that each subproblem in the proposed algorithm can be solved in closed-form and it can be shown that the convergence to the set of KKT solutions is guaranteed. Furthermore, we provided a detailed complexity analysis of the proposed algorithm along with its extension to the subarray structure. The simulation results validated the effectiveness of the proposed PDD-based algorithm and the necessity of using a robust formalism in the presence of imperfect CSI.

APPENDIX

A. Proof of Equivalence Between Problem (15) and Problem (22)

Since $\{u_k^{(n)}\}$, $\{w_k^{(n)}\}$, $\{\mathbf{U}_{\text{BB}}^{(n)}\}$ and $\{\mathbf{W}_b^{(n)}\}$ only appear in the objective function of problem (22), therefore, we can see that problem (22) is equivalent to problem (49), shown at the top of the next page. As a result, by sequentially solving problem (49) w.r.t $\{u_k^{(n)}, w_k^{(n)}\}$ and $\{\mathbf{U}_{\text{BB}}^{(n)}, \mathbf{W}_b^{(n)}\}$, we will obtain an equivalent problem which is only associated with $\{\mathcal{X}^{(n)} \setminus \mathbf{U}_{\text{BB}}^{(n)}\}$. Specifically, solving the innermost minimization problem for $\{\mathbf{U}_{\text{BB}}^{(n)}, \mathbf{W}_b^{(n)}\}$, we can obtain

$$\mathbf{U}_{\text{BB}}^{(n)} = (\mathbf{A}^{(n)})^{-1} \mathbf{B}^{(n)}, \quad (50)$$

$$\mathbf{W}_b^{(n)} = (\mathbf{E}_b^{(n)})^{-1}, \quad (51)$$

where $\mathbf{A}^{(n)} \triangleq (\Upsilon_{\text{BS}}^{(n)} + \mathbf{U}_{\text{RF}}^{(n)H} \hat{\mathbf{G}}^{(n,n)} \mathbf{P}^{(n)} \hat{\mathbf{G}}^{(n,n)H} \mathbf{U}_{\text{RF}}^{(n)})$ and $\mathbf{B}^{(n)} \triangleq \mathbf{U}_{\text{RF}}^{(n)H} \hat{\mathbf{G}}^{(n,n)} \mathbf{P}^{(n)\frac{1}{2}}$. Similarly, we can obtain the optimal solutions of $\{u_k^{(n)}, w_k^{(n)}\}$ by

$$u_k^{(n)} = \frac{\hat{\mathbf{h}}_k^{(n,n)H} \mathbf{V}_{\text{RF}}^{(n)} \mathbf{V}_{\text{BB},k}^{(n)}}{\Upsilon_k^{(n)} + |\hat{\mathbf{h}}_k^{(n,n)H} \mathbf{V}_{\text{RF}}^{(n)} \mathbf{V}_{\text{BB},k}^{(n)}|^2}, \quad (52)$$

$$w_k^{(n)} = (e_k^{(n)})^{-1}. \quad (53)$$

$$\min_{\{\boldsymbol{\chi}^{(n)} \setminus \mathbf{U}_{\text{BB}}^{(n)}\}} \left(\min_{\{u_k^{(n)}, w_k^{(n)}\}} \left(\min_{\{\mathbf{U}_{\text{BB}}^{(n)}, \mathbf{W}_b^{(n)}\}} \left(\sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} h^{(n,k)}(u_k^{(n)}, w_k^{(n)}, \boldsymbol{\chi}^{(n)}) + \sum_{n \in \mathcal{N}} g^{(n)}(\mathbf{W}_b^{(n)}, \boldsymbol{\chi}^{(n)}) \right) \right) \right) \quad (49)$$

s.t. (15b), (15c) and (15d).

By substituting (50)-(53) back into (49), it yields the following equivalent problem:

$$\max_{\{\boldsymbol{\chi}^{(n)} \setminus \mathbf{U}_{\text{BB}}^{(n)}\}} \sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} \log \left(1 - \frac{|\hat{\mathbf{h}}_k^{(n,n)H} \mathbf{V}_{\text{RF}}^{(n)} \mathbf{V}_{\text{BB},k}^{(n)}|^2}{|\hat{\mathbf{h}}_k^{(n,n)H} \mathbf{V}_{\text{RF}}^{(n)} \mathbf{V}_{\text{BB},k}^{(n)}|^2 + \Upsilon_k^{(n)}} \right)^{-1} + \sum_{n \in \mathcal{N}} \log \det |(\mathbf{I} - \mathbf{B}^{(n)H} (\mathbf{A}^{(n)})^{-1} \mathbf{B}^{(n)})^{-1}| \quad (54)$$

s.t. (15b), (15c) and (15d).

Then, by resorting to the Woodbury identity and $\det(\mathbf{I} + \mathbf{A}_1 \mathbf{A}_2) = \det(\mathbf{I} + \mathbf{A}_2 \mathbf{A}_1)$, it can be seen that (54) is equivalent to (15), therefore, it follows that (22) and (15) are equivalent in the sense that they have the same global optimal solution. Furthermore, by following [37, Theorem 3] and [52, Lemma 4.1], it can be readily proved that if $\{\boldsymbol{\chi}^{(n)}, u_k^{(n)}, w_k^{(n)}, \mathbf{W}_b^{(n)}\}$ is a KKT point of problem (22), then $\{\boldsymbol{\chi}^{(n)}\}$ is a KKT point of problem (15). This completes the proof.

B. Optimal Solution to Problem (37)

In order to simplify the notations, we define $a \triangleq q_{d,j}^{(n)}$, $b \triangleq [\mathbf{W}_b^{(n)} \mathbf{U}^{(n)H} \hat{\mathbf{G}}^{(n,n)} + \hat{\mathbf{G}}^{(n,n)H} \mathbf{U}^{(n)} \mathbf{W}_b^{(n)}]_{jj}$, $x \triangleq q_j^{(n)}$ and the subscript (superscript) j (n) is neglected here without loss of generality. Then, problem (37) can be rewritten as

$$\min_x ax^2 + bx \quad (55)$$

s.t. $x^2 \leq P_u$.

The Lagrangian function of problem (55) can be expressed as

$$\mathcal{L}(x, \lambda) = ax^2 + bx + \lambda(x^2 - P_u), \quad (56)$$

where λ denotes the dual variable. Resorting to the first-order optimality condition $\nabla_x \mathcal{L}((x, \lambda)) = 0$, we have $x^{\text{opt}} = -0.5(b + \lambda)/a$. According to the complementary slackness condition [53], we have $(b + \lambda^{\text{opt}})^2 = 4a^2 P_u$, therefore, the optimal solution can be obtained by

$$x^{\text{opt}} = \begin{cases} -\frac{b}{2a}, & \text{if } \frac{b^2}{4a^2} \leq P_u, \\ -\frac{b + \lambda^{\text{opt}}}{2a}, & \text{otherwise,} \end{cases} \quad (57)$$

where $\lambda^{\text{opt}} = -b + \sqrt{\frac{P_u}{4a^2}}$.

C. MFCQ Verification for Problem (23)

According to [47], the MFCQ for problem (16) is equivalent to verify that the following two conditions are satisfied for any feasible solution $\hat{\mathbf{x}}$: 1) the equality constraint gradients are linearly independent; 2) there exists a vector \mathbf{d} such that $\nabla \mathbf{h}(\hat{\mathbf{x}}) \mathbf{d} = \mathbf{0}$, $\nabla \mathbf{g}_{i,j}(\hat{\mathbf{x}}_i)^T \mathbf{d} < 0$, $\forall i, \forall j \in \mathcal{A}(\hat{\mathbf{x}}_i)$ are satisfied, where $\nabla \mathbf{h}(\hat{\mathbf{x}})$ is the Jacobian matrix of $\mathbf{h}(\mathbf{x})$, $\nabla \mathbf{g}_{i,j}(\hat{\mathbf{x}}_i)$ denotes the gradient of $\mathbf{g}_{i,j}(\mathbf{x}_i)$ and $\mathbf{g}_{i,j}(\mathbf{x}_i)$ denotes the j -th component of $\mathbf{g}_i(\mathbf{x}_i)$, $\mathcal{A}(\hat{\mathbf{x}}_i)$ is the set that contains the indices of active inequality constraints. In the following, we show that they are satisfied for problem (23).

The equality constraints of problem (23) are written as follows:

$$\begin{aligned} \Theta_1^{(n)} &\triangleq \mathbf{V}^{(n)} - \tilde{\mathbf{V}}_{\text{RF}}^{(n)} \mathbf{V}_{\text{BB}}^{(n)} = \mathbf{0}, \\ \Theta_2^{(n)} &\triangleq \mathbf{U}^{(n)} - \tilde{\mathbf{U}}_{\text{RF}}^{(n)} \mathbf{U}_{\text{BB}}^{(n)} = \mathbf{0}, \\ \Theta_3^{(n)} &\triangleq \mathbf{V}^{(n)} - \tilde{\mathbf{V}}^{(n)} = \mathbf{0}, \\ \Theta_4^{(n)} &\triangleq \mathbf{V}_{\text{RF}}^{(n)} - \tilde{\mathbf{V}}_{\text{RF}}^{(n)} = \mathbf{0}, \\ \Theta_5^{(n)} &\triangleq \mathbf{U}_{\text{RF}}^{(n)} - \tilde{\mathbf{U}}_{\text{RF}}^{(n)} = \mathbf{0}. \end{aligned} \quad (58)$$

Since constraints $\Theta_1^{(n)}$ and $\Theta_2^{(n)}$ do not contain variable $\tilde{\mathbf{V}}^{(n)}$ while constraint $\Theta_3^{(n)}$ does, thus the gradient of the components of $\Theta_3^{(n)}$ is independent with the other equality constraints. By a similar argument, we can see that $\{\Theta_i^{(n)}\}, i \in \{1, \dots, 5\}$ are linearly independent with each other since they all contain variables that the others do not have. Therefore, we can conclude that the equality constraint gradients of problem (23) are linearly independent.

Then, we are left to show that there exists $\{\Delta_{\mathbf{V}_{\text{RF}}^{(n)}}, \Delta_{\tilde{\mathbf{V}}_{\text{RF}}^{(n)}}, \Delta_{\mathbf{V}_{\text{BB}}^{(n)}}, \Delta_{\mathbf{U}_{\text{RF}}^{(n)}}, \Delta_{\tilde{\mathbf{U}}_{\text{RF}}^{(n)}}, \Delta_{\mathbf{U}_{\text{BB}}^{(n)}}, \Delta_{\mathbf{P}^{(n)}}, \Delta_{\mathbf{V}^{(n)}}, \Delta_{\mathbf{U}^{(n)}}, \Delta_{\tilde{\mathbf{V}}^{(n)}}\}$, such that the following conditions are satisfied [47, Remark C.1]:

$$\Delta_{P_j^{(n)}} < 0, \quad \forall j, n, \quad (59a)$$

$$\Re\{\text{Tr}(\tilde{\mathbf{V}}^{(n)} \Delta_{\tilde{\mathbf{V}}^{(n)}}^H)\} < 0, \quad \forall n, \quad (59b)$$

$$\Delta_{\mathbf{V}^{(n)}} - \Delta_{\tilde{\mathbf{V}}_{\text{RF}}^{(n)}} \mathbf{V}_{\text{BB}}^{(n)} - \tilde{\mathbf{V}}_{\text{RF}}^{(n)} \Delta_{\mathbf{V}_{\text{BB}}^{(n)}} = \mathbf{0}, \quad \forall n, \quad (59c)$$

$$\Delta_{\mathbf{U}^{(n)}} - \Delta_{\tilde{\mathbf{U}}_{\text{RF}}^{(n)}} \mathbf{U}_{\text{BB}}^{(n)} - \tilde{\mathbf{U}}_{\text{RF}}^{(n)} \Delta_{\mathbf{U}_{\text{BB}}^{(n)}} = \mathbf{0}, \quad \forall n, \quad (59d)$$

$$\Delta_{\mathbf{V}^{(n)}} - \Delta_{\tilde{\mathbf{V}}^{(n)}} = \mathbf{0}, \quad \forall n, \quad (59e)$$

$$\Delta_{\mathbf{V}_{\text{RF}}^{(n)}} - \Delta_{\tilde{\mathbf{V}}_{\text{RF}}^{(n)}} = \mathbf{0}, \quad \forall n, \quad (59f)$$

$$\Delta_{\mathbf{U}_{\text{RF}}^{(n)}} - \Delta_{\tilde{\mathbf{U}}_{\text{RF}}^{(n)}} = \mathbf{0}, \quad \forall n, \quad (59g)$$

where (59c)-(59g) are obtained by using the first-order approximations of the equality constraints. By choosing $\Delta_{\mathbf{V}_{\text{RF}}^{(n)}} = \Delta_{\tilde{\mathbf{V}}_{\text{RF}}^{(n)}} = -\mathbf{V}_{\text{RF}}^{(n)}$, $\Delta_{\mathbf{V}_{\text{BB}}^{(n)}} = -\mathbf{V}_{\text{BB}}^{(n)}$, $\Delta_{\mathbf{U}_{\text{RF}}^{(n)}} = \Delta_{\tilde{\mathbf{U}}_{\text{RF}}^{(n)}} = -\mathbf{U}_{\text{RF}}^{(n)}$, $\Delta_{\mathbf{U}_{\text{BB}}^{(n)}} = -\mathbf{U}_{\text{BB}}^{(n)}$, $\Delta_{\mathbf{P}^{(n)}} = -\mathbf{P}^{(n)}$, $\Delta_{\mathbf{V}^{(n)}} = -2\mathbf{V}^{(n)}$, $\Delta_{\mathbf{U}^{(n)}} = -2\mathbf{U}^{(n)}$, $\Delta_{\tilde{\mathbf{V}}^{(n)}} = -2\mathbf{V}^{(n)}$, (59) is satisfied with $\mathbf{P}^{(n)} \neq \mathbf{0}$ and $\mathbf{V}^{(n)} \neq \mathbf{0}$, $\forall n$. Note that we here consider only the case when (23b) and (23c) are all active. Other cases can be similarly treated, thus we omit the details here. This completes the verification.

REFERENCES

- [1] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.
- [2] C. Liu, M. Li, S. V. Hanly, I. B. Collings, and P. Whiting, "Millimeter wave beam alignment: Large deviations analysis and design insights," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1619–1631, Jul. 2017.
- [3] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda, "Hybrid beamforming for massive MIMO: A survey," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, Sep. 2017.

- [4] C. Liu, M. Li, S. V. Hanly, P. Whiting, and I. B. Collings, "Millimeter-wave small cells: Base station discovery, beam alignment, and system design challenges," *IEEE Wireless Commun.*, vol. 25, no. 4, pp. 40–46, Aug. 2018.
- [5] M. Li, C. Liu, S. V. Hanly, I. B. Collings, and P. Whiting, "Explore and eliminate: Optimized two-stage search for millimeter-wave beam alignment," *IEEE Trans. Wireless Commun.*, vol. 18, no. 9, pp. 4379–4393, Sep. 2019.
- [6] X. Yu, J. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, Apr. 2016.
- [7] 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.
- [8] R. Rajashekar and L. Hanzo, "Iterative matrix decomposition aided block diagonalization for mm-wave multiuser MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1372–1384, Mar. 2017.
- [9] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
- [10] C. Chen, "An iterative hybrid transceiver design algorithm for millimeter wave MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 4, no. 3, pp. 285–288, Jun. 2015.
- [11] W. Ni, X. Dong, and W. Lu, "Near-optimal hybrid processing for massive MIMO systems via matrix decomposition," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 3922–3933, Aug. 2017.
- [12] F. Sotrabadi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.
- [13] Z. Xiao, T. He, P. Xia, and X. Xia, "Hierarchical codebook design for beamforming training in millimeter-wave communication," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3380–3392, May 2016.
- [14] 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, Sep. 2017.
- [15] 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 mmWave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 2085–2096, Sep. 2017.
- [16] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 653–656, Dec. 2014.
- [17] J. Zhang, M. Haardt, I. Soloveyichik, and A. Wiesel, "A channel matching based hybrid analog-digital strategy for massive multi-user MIMO downlink systems," in *IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Jul. 2016, pp. 1–5.
- [18] W. Ni and X. Dong, "Hybrid block diagonalization for massive multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 201–211, Jan. 2016.
- [19] T. Riihonen, S. Werner, R. Wichman, and E. Zaccarias B., "On the feasibility of full-duplex relaying in the presence of loop interference," in *IEEE 10th Workshop on Signal Processing Advances in Wireless Communications*, Jun. 2009, pp. 275–279.
- [20] T. Riihonen, S. Werner, and R. Wichman, "Hybrid full-duplex/half-duplex relaying with transmit power adaptation," *IEEE Trans. Wireless Commun.*, vol. 10, no. 9, pp. 3074–3085, Sep. 2011.
- [21] P. Aquilina, A. C. Cirik, and T. Ratnarajah, "Weighted sum rate maximization in full-duplex multi-user multi-cell MIMO networks," *IEEE Trans. Commun.*, vol. 65, no. 4, pp. 1590–1608, Apr. 2017.
- [22] M. M. Zhao, Q. Shi, Y. Cai, and M. Zhao, "Joint transceiver design for full-duplex cloud radio access networks with SWIPT," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5644–5658, Sep. 2017.
- [23] X. Liu, Z. Xiao, L. Bai, J. Choi, P. Xia, and X.-G. Xia, "Beamforming based full-duplex for millimeter-wave communication," *Sensors*, vol. 17, no. 7, pp. 1–22, Jul. 2016.
- [24] Z. Xiao, P. Xia, and X. Xia, "Full-duplex millimeter-wave communication," *IEEE Wireless Commun.*, vol. 24, no. 6, pp. 136–143, Dec. 2017.
- [25] T. Dinc and H. Krishnaswamy, "Millimeter-wave full-duplex wireless: Applications, antenna interfaces and systems," in *IEEE Custom Integrated Circuits Conference (CICC)*, Apr. 2017, pp. 1–8.
- [26] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4296–4307, Dec. 2012.
- [27] M. Jain *et al.*, "Practical, real-time, full duplex wireless," in *Proceedings of the 17th Annual International Conference on Mobile Computing and Networking (MOBICOM)*, Sep. 2011, pp. 301–312.
- [28] Z. Wei, X. Zhu, S. Sun, Y. Huang, A. Al-Tahmeesschi, and Y. Jiang, "Energy-efficiency of millimeter-wave full-duplex relaying systems: Challenges and solutions," *IEEE Access*, vol. 4, pp. 4848–4860, 2016.
- [29] H. Abbas and K. Hamdi, "Full duplex relay in millimeter wave backhaul links," in *IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2016, pp. 1–6.
- [30] B. Ma, H. Shah-Mansouri, and V. W. S. Wong, "Full-duplex relaying for D2D communication in millimeter wave-based 5G networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4417–4431, Jul. 2018.
- [31] G. Yang, M. Xiao, H. Al-Zubaidy, Y. Huang, and J. Gross, "Analysis of millimeter-wave multi-hop networks with full-duplex buffered relays," *IEEE/ACM Trans. Netw.*, vol. 26, no. 1, pp. 576–590, Feb. 2018.
- [32] W. Ding, Y. Niu, H. Wu, Y. Li, and Z. Zhong, "Qos-aware full-duplex concurrent scheduling for millimeter wave wireless backhaul networks," *IEEE Access*, vol. 6, pp. 25 313–25 322, 2018.
- [33] K. Satyanarayana, M. El-Hajjar, P. Kuo, A. Mourad, and L. Hanzo, "Hybrid beamforming design for full-duplex millimeter wave communication," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1394–1404, Feb. 2019.
- [34] Z. Luo, H. Liu, Y. Li, H. Wang, and L. Zhang, "Robust hybrid transceiver design for AF relaying in millimeter wave systems under imperfect CSI," *IEEE Access*, vol. 6, pp. 29 739–29 746, 2018.
- [35] D. Jagyasi and P. Ubaidulla, "In-band full-duplex relay-assisted millimeter-wave system design," *IEEE Access*, vol. 7, pp. 2291–2304, 2019.
- [36] Y. Cai, Y. Xu, Q. Shi, B. Champagne, and L. Hanzo, "Robust joint hybrid transceiver design for millimeter wave full-duplex MIMO relay systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1199–1215, Feb. 2019.
- [37] Q. Shi, M. Razaviyayn, Z. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.
- [38] M. Hong, M. Razaviyayn, Z. Luo, and J. Pang, "A unified algorithmic framework for block-structured optimization involving big data: With applications in machine learning and signal processing," *IEEE Signal Process. Mag.*, vol. 33, no. 1, pp. 57–77, Jan. 2016.
- [39] Y. Choi and H. Shirani-Mehr, "Simultaneous transmission and reception: Algorithm, design and system level performance," *IEEE Trans. Wireless Commun.*, vol. 12, no. 12, pp. 5992–6010, Dec. 2013.
- [40] J. Dai, A. Liu, and V. K. N. Lau, "FDD massive MIMO channel estimation with arbitrary 2D-array geometry," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2584–2599, May 2018.
- [41] L. Chen, A. Liu, and X. Yuan, "Structured turbo compressed sensing for massive MIMO channel estimation using a markov prior," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4635–4639, May 2018.
- [42] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.
- [43] S. Lakshminarayana, M. Assaad, and M. Debbah, "Coordinated multicell beamforming for massive MIMO: A random matrix approach," *IEEE Trans. Inf. Theory*, vol. 61, no. 6, pp. 3387–3412, Jun. 2015.
- [44] M. Ding and S. D. 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.
- [45] L. Musavian, M. R. Nakhai, M. Dohler, and A. H. Aghvami, "Effect of channel uncertainty on the mutual information of MIMO fading channels," *IEEE Trans. Veh. Technol.*, vol. 56, no. 5, pp. 2798–2806, Sep. 2007.
- [46] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.
- [47] Q. Shi and M. Hong, "Penalty dual decomposition method for non-smooth nonconvex optimization Part I & Part II," *ArXiv e-prints 1712.04767*, 2017.
- [48] M. M. Zhao, Q. Shi, Y. Cai, M. Zhao, and Q. Yu, "Decoding binary linear codes using penalty dual decomposition method," *IEEE Commun. Lett.*, vol. 23, no. 6, pp. 958–962, Jun. 2019.
- [49] M. M. Zhao, Q. Shi, Y. Cai, M. Zhao, and Y. Li, "Distributed penalty dual decomposition algorithm for optimal power flow in radial networks," *IEEE Trans. Power Syst.*, DOI10.1109/TPWRS.2019.2952433, 2019.
- [50] Q. Shi and M. Hong, "Penalty dual decomposition method with application in signal processing," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2017, pp. 4059–4063.

- [51] J. A. Zhang, X. Huang, V. Dyadyuk, and Y. J. Guo, "Massive hybrid antenna array for millimeter-wave cellular communications," *IEEE Wireless Commun.*, vol. 22, no. 1, pp. 79–87, Feb. 2015.
- [52] Q. Shi, W. Xu, J. Wu, E. Song, and Y. Wang, "Secure beamforming for MIMO broadcasting with wireless information and power transfer," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2841–2853, May 2015.
- [53] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.