Majorization-Minimization Aided Hybrid Transceivers for MIMO Interference Channels

Shiqi Gong, Chengwen Xing, Vincent K. N. Lau, *Fellow, IEEE*, Sheng Chen, *Fellow, IEEE*, and Lajos Hanzo, *Fellow, IEEE*

Abstract—The potential of deploying large-scale antenna arrays in future wireless systems has stimulated extensive research on hybrid transceiver designs aiming to approximate the optimal fully-digital schemes with much reduced hardware cost and signal processing complexity. Generally, this hybrid transceiver structure requires a joint design of analog and digital processing to enable both beamsteering and spatial multiplexing gains. In this paper, we develop various weighted mean-square-error minimization (WMMSE) based hybrid transceiver designs for K-user multiple-input multiple-output (MIMO) interference systems, which are applicable to both millimeter wave (mmWave) channels and Rayleigh fading channels. Firstly, a heuristic joint design of hybrid precoder and combiner using alternating optimization is proposed, in which the majorization-minimization (MM) method is utilized to design the analog precoder and combiner under unit-modulus constraints. It is demonstrated that this scheme achieves comparable performance to the fully-digital WMMSE solution. To further reduce the computational complexity, a phase projection based two-stage scheme is proposed to decouple the designs of the analog and digital precoder/combiner. Secondly, inspired by the fully-digital solutions based on the block-diagonalization zero-forcing (BD-ZF) and signal-to-leakage-plus-noise ratio (SLNR) criteria, the low-complexity MM-based BD-ZF and SLNR hybrid designs are proposed, respectively, for approximating the corresponding fully-digital solutions. Thirdly, the partially-connected hybrid structure conceived for reducing system hardware cost and power consumption is considered, for which the MM-based alternating optimization algorithm still works. Our numerical results characterize the sum rate performance of all proposed hybrid designs in comparison to the existing benchmarks.

Index Terms—Hybrid transceiver designs, WMMSE, majorization-minimization, MIMO interference channels.

I. INTRODUCTION

The large-scale antenna array offers a promising technology in future wireless systems to provide ultra high data rate for bandwidth-hungry applications and the large degrees of freedom (DoFs) for eliminating the random effect of wireless fading channels [1]–[3]. However, the hardware cost and implementation complexity of deploying a large number of antenna elements by the traditional digital signal processing are huge, because each antenna requires a dedicated radio

frequency (RF) chain [4], [5]. As an alternative cost-effective solution, the hybrid transceiver structure with much fewer RF chains than the number of antennas has attracted extensive attention recently, of which the signal processing chain consists of the high-dimensional analog RF precoding/combining for providing the beamsteering gain, followed by the low-dimensional digital baseband precoding/combing mainly for reaping spatial multiplexing gain [6], [7].

For the hybrid transceiver structure, the analog RF processing can be implemented using phase shifters [8], switches [9] and/or lens [10], among which the phase shifter based analog precoding/combining has been widely investigated [11]–[18]. Phase shifters can be used to steer transmit and receive beams towards the desired direction by adjusting the phase of RF signals, and thus typically impose constant-modulus constraints on analog precoder and combiner, which makes hybrid transceiver designs more complicated and challenging. It has been revealed that once the number of RF chains reaches twice that of data streams, implying that the number of phase shifters is doubled, the hybrid structure can perfectly realize the optimal fully-digital structure [12]. However, the application with abundant phase shifters is also impractical due to high hardware cost and power consumption. To alleviate this issue, the partially-connected hybrid structure has been proposed for enabling energy-efficient communications at the expense of some performance loss compared to the fully-digital structure [13]–[15].

Hybrid transceivers are applicable not only to mmWave communications but also in other lower frequency range [11], [16]. Moreover, the criteria of hybrid designs are diverse, e.g., mean squared error (MSE), capacity and bit error rate (BER). Various hybrid transceiver designs have been conceived for point-to-point MIMO systems [7], [18]–[23] and multiuser MIMO systems [16], [24]–[28]. The motivation of these designs is to leverage the underlying hybrid structure to achieve the comparable performance to the optimal (near-optimal) fully-digital solution. To this end, existing hybrid designs are mainly classified into two categories.

One category jointly designs hybrid precoder and combiner to approach the fully-digital performance. For example, by exploiting the sparsity of mmWave channels, the orthogonal matching pursuit (OMP) algorithm was used to jointly design hybrid precoder and combiner to approximate the optimal fully-digital solution [7]. Using matrix-monotonic optimization [18], the optimal unconstrained structures of analog precoder and combiner under various design criteria can be proved to be unitary matching with channel. Some heuristic joint hybrid transceiver designs via alternating optimization were also investigated [19]–[23]. Specifically, to approximate the optimal fully-digital solution, an alternating

minimization method was proposed for hybrid designs based on manifold optimization [19] and local approximation of phase increment [20], [23], respectively. In addition, joint hybrid designs were studied in multiuser scenarios using the minimum MSE (MMSE), weighted MMSE (WMMSE) and block-diagonalization zero-forcing (BD-ZF) fully-digital solutions [24]–[26] For example, in [25] and [26], the OMP algorithm was utilized to jointly construct the hybrid WMMSE precoder and combiner for achieving the performance close to the WMMSE and BD-ZF fully-digital solutions, respectively. However, such approaches generally require the fully-digital precoder to have a closed-form solution, and its applicability in more general scenarios may be limited.

The other category is the two-stage hybrid transceiver design widely used in multiuser MIMO scenarios. In this scheme, the analog precoder and combiner are firstly designed by directly optimizing some performance criterion, such as the effective array gain. Then the digital precoder and combiner are optimized to further improve system performance by eliminating inter-user inference [16], [27]–[29]. For example, in [16], the equal gain transmission (EGT) based analog precoder and the discrete Fourier transform (DFT) codebook based analog combiner for each user were proposed to achieve large array gain. To achieve low channel training and feedback overhead, the two-stage hybrid design [27] chooses each user's analog precoder and combiner from the quantized codebooks to maximize effective channel gain. All the above analog processing schemes can be combined with the low-complexity BD-ZF digital processing [30] to cancel inter-user interference. Although this BD-ZF scheme is easy to implement, it does not consider the influence of noise in the digital precoder design and thus performs poorly at low signal-to-noise ratio (SNR) regime. This fact motivates us to consider an effective digital processing based on the SLNR criterion of [31]. The SLNR-maximization digital processing is more desirable than the BD-ZF criterion in some scenarios with fewer DoFs, i.e., MIMO interference channels [32]. This two-stage scheme can also be extended to the mixed timescale hybrid precoder optimization [33], [34] in which the analog and digital precoders are adaptive to channel statistics and real-time channel state information (CSI), respectively.

In this paper, we consider challenging MIMO interference channels with very few DoFs and develop various hybrid transceiver designs based on the MM method. Since the MM method guarantees stationary convergence and has the desired closed-form solution of each subproblem, it offers an effective tool to address the nonconvex constant-modulus constraints on analog precoder and combiner [35], [38]. Specifically, we propose the MM-based alternating optimization, decoupled two-stage scheme and various low-complexity schemes for hybrid transceiver designs in both mmWave and lower-frequency Rayleigh MIMO interference channels. Additionally, perfect CSI and analog processing with infinite resolution are utilized to provide a theoretical performance upper-bound for practical implementation of all the proposed schemes. Our contributions are contrasted to the associated technical challenges as follows:

1) **Joint Hybrid Transceiver Design Bypassing the Optimal Fully-Digital Solution**. For the K-user MIMO interference channel, the joint hybrid WMMSE transceiver design bypassing the near-optimal fully-digital WMMSE solution is studied. This joint design is challenging since the coupled variables and the unit-modulus constraint imposed on the analog precoder and combiner lead to nonconvex and NP-hard optimization. To tackle this challenge, MM-based alternating optimization considering a practical property of large-scale MIMO is proposed, which guarantees to converge. To further reduce the computational complexity, we also study another phase projection (PP) based two-stage scheme relying on the decoupled design of the analog and digital precoder and combiner.

2) **Low-Complexity Separate Hybrid Transceiver Designs**. Since the suboptimal closed-form fully-digital precoders designed for each transmit-receive pair can be obtained based on the BD-ZF and SLNR maximization (SLNR-Max) criteria, the proposed low-complexity hybrid transceiver designs focus on approximating the BD-ZF and SLNR-Max based fully-digital precoders, which also lead to nonconvex optimization. In fact, both these low-complexity designs rely on multiple separate hybrid transceiver designs for all transmit-receive pairs, each of which consists of two separate stages. To address this non-convexity, first an iterative PP (iterative-PP) based hybrid precoder is designed. Then the corresponding hybrid MMSE combiner is optimized using MM-based alternating optimization.

3) **Low-Cost Joint Hybrid Transceiver Design**. In order to further reduce both the hardware cost and power consumption, we consider the partially-connected hybrid structure, in which each RF chain at the transmitter/receiver is connected to a single non-overlapped subarray. In this context, the MM-based alternating optimization still works and can converge to the stationary solutions for our joint hybrid WMMSE transceiver design.

The rest of this paper is organized as follows. Section II introduces the system model and the problem formulation. Section III proposes the MM-based alternating optimization algorithm and the low-complexity two-stage hybrid design for maximizing the sum rate of the considered system, respectively. Section IV mainly introduces the hybrid transceiver designs from the perspective of low complexity and low cost. Section V illustrates the convergence and complexity of the proposed MM-based alternating optimization algorithm. Our numerical results are presented in Section VI to demonstrate the excellent performance of the proposed hybrid designs. Finally, Section VII concludes this paper.

Notations: The bold-faced lower-case and upper-case letters stand for vectors and matrices, respectively. The transpose, conjugate, Hermitian and inverse operators are denoted by $(\cdot)^{\mathrm{T}}$, $(\cdot)^{*}$, $(\cdot)^{\mathrm{H}}$ and $(\cdot)^{-1}$, respectively, while $\mathrm{Tr}(\boldsymbol{A})$ and $\det(\boldsymbol{A})$ denote the trace and determinant of $\boldsymbol{A}$, respectively. $\boldsymbol{I}_n$, $\boldsymbol{0}_{n \times m}$ and $\boldsymbol{1}_n$ are the $n \times n$ identity matrix, the $n \times m$ zero matrix and the n-dimensional vector with all elements being one, respectively. The block-diagonal matrix with diagonal

elements $\boldsymbol{A}_1, \cdots, \boldsymbol{A}_N$ is denoted by BLKdiag$[\boldsymbol{A}_1, \cdots, \boldsymbol{A}_N]$. Particularly, it is reduced to diag$[a_1, \cdots, a_N]$ when scalar diagonal elements are considered . $[\boldsymbol{A}]_{n,m}$ denotes the (n, m)th (the nth row and mth column) element of $\boldsymbol{A}$, and $\boldsymbol{A}(q_1 : q_2, l_1 : l_2)$ denotes the sub-matrix consisting of the q_1 to q_2 rows and l_1 to l_2 columns of $\boldsymbol{A}$, while $\boldsymbol{A}(:, l_1 : l_2)$ is the sub-matrix consisting of the l_1 to l_2 columns of $\boldsymbol{A}$. The nth element of $\boldsymbol{a}$ is denoted by $[\boldsymbol{a}]_n$, and $\boldsymbol{a}(n : m)$ is the sub-vector consists of the nth to mth elements of $\boldsymbol{a}$. $\boldsymbol{A} \succ \boldsymbol{0} \ (\succeq \boldsymbol{0})$ means that $\boldsymbol{A}$ is positive definite (semi-definite), and $\lambda_{\max}(\boldsymbol{A})$ is the maximum eigenvalue of $\boldsymbol{A}$, while $e^{\mathrm{j}\arg(\cdot)}$ denotes the phase extraction operation in an element-wise manner. The rank of $\boldsymbol{A}$ is denoted by rank$(\boldsymbol{A})$. The modulus operator denoted by $|\cdot|$, $\|\cdot\|$ is the Euclidean distance, and $\|\cdot\|_F$ is the matrix Frobenius norm, while $\mathbb{E}[\cdot]$ is the expectation operator and vec$(\cdot)$ denotes the vectorization of a matrix. $\Re\{\cdot\}$ is the real part operator and $\otimes$ is the Kronecker product operator, while $[a]^+ = \max\{a, 0\}$. The words 'independent and identically distributed' and 'with respect to' are abbreviated as 'i.i.d.' and 'w.r.t.', respectively.

II. System Model

A. K-user MIMO Interference Channel

As shown in Fig. 1, we consider a K-user MIMO interference channel, where all K transmitters and receivers are equipped with hybrid MIMO processor for dealing with multiple data streams. Specifically, the kth transmitter equipped with N_{t_k} antennas and $N_{t_k}^{RF}$ RF chains sends N_{s_k} data streams to the corresponding receiver equipped with N_{r_k} antennas and $N_{r_k}^{RF}$ RF chains, where $N_{s_k} \leq N_{t_k}^{RF} \leq N_{t_k}$ and $N_{s_k} \leq N_{r_k}^{RF} \leq N_{r_k}$, $\forall k$. The hybrid MIMO processor at the kth transmitter enables the digital baseband precoder $\boldsymbol{F}_{D_k} \in \mathbb{C}^{N_{t_k}^{RF} \times N_{s_k}}$, followed by the analog RF precoder $\boldsymbol{F}_{A_k} \in \mathbb{C}^{N_{t_k} \times N_{t_k}^{RF}}$. Similarly, the hybrid MIMO processor at the kth receiver consists of an analog RF combiner $\boldsymbol{G}_{A_k} \in \mathbb{C}^{N_{r_k} \times N_{r_k}^{RF}}$, followed by a digital combiner $\boldsymbol{G}_{D_k} \in \mathbb{C}^{N_{r_k}^{RF} \times N_{s_k}}$. Both $\boldsymbol{F}_{A_k}$ and $\boldsymbol{G}_{A_k}$ are realized using analog phase shifters with constant modulus, i.e. $|[\boldsymbol{F}_{A_k}]_{n,m}| = 1$ and $|[\boldsymbol{G}_{A_k}]_{n,m}| = 1$, $\forall n, m$. The transmitted signal by the kth transmitter is given by $\boldsymbol{x}_k = \boldsymbol{F}_{A_k} \boldsymbol{F}_{D_k} \boldsymbol{s}_k$, where $\boldsymbol{s}_k \in \mathbb{C}^{N_{s_k}}$ denotes the Gaussian encoded information symbols satisfying $\mathbb{E}[\boldsymbol{s}_k \boldsymbol{s}_k^{\mathrm{H}}] = \boldsymbol{I}_{N_{s_k}}$ and $\|\boldsymbol{F}_{A_k} \boldsymbol{F}_{D_k}\|_F^2 \leq P_k$ with P_k being the maximum transmission power. Under the assumption of quasi-static block-fading MIMO channel, the received signal at the kth receiver is written as

$$\boldsymbol{y}_k = \boldsymbol{H}_{k,k} \boldsymbol{F}_{A_k} \boldsymbol{F}_{D_k} \boldsymbol{s}_k + \sum_{i \neq k} \boldsymbol{H}_{k,i} \boldsymbol{F}_{A_i} \boldsymbol{F}_{D_i} \boldsymbol{s}_i + \boldsymbol{n}_k, \quad (1)$$

where $\boldsymbol{H}_{k,i} \in \mathbb{C}^{N_{r_k} \times N_{t_i}}$ denotes the wireless channel between the ith transmitter and kth receiver, and $\boldsymbol{n}_k \sim \mathcal{CN}(\boldsymbol{0}, \sigma_{n_k}^2 \boldsymbol{I}_{N_{r_k}})$ is the additive Gaussian noise at the kth receiver, which has zero mean vector and covariance matrix $\sigma_{n_k}^2 \boldsymbol{I}_{N_{r_k}}$. Then the hybrid analog-digital combiner at the kth receiver, i.e. $\boldsymbol{G}_k^{\mathrm{H}} = \boldsymbol{G}_{D_k}^{\mathrm{H}} \boldsymbol{G}_{A_k}^{\mathrm{H}}$, is applied to $\boldsymbol{y}_k$ to obtain the desired output as

$$\begin{aligned} \widehat{\boldsymbol{s}}_k = &\ \boldsymbol{G}_{D_k}^{\mathrm{H}} \boldsymbol{G}_{A_k}^{\mathrm{H}} \boldsymbol{H}_{k,k} \boldsymbol{F}_{A_k} \boldsymbol{F}_{D_k} \boldsymbol{s}_k \\ &+ \boldsymbol{G}_{D_k}^{\mathrm{H}} \boldsymbol{G}_{A_k}^{\mathrm{H}} \sum_{i \neq k} \boldsymbol{H}_{k,i} \boldsymbol{F}_{A_i} \boldsymbol{F}_{D_i} \boldsymbol{s}_i + \boldsymbol{G}_{D_k}^{\mathrm{H}} \boldsymbol{G}_{A_k}^{\mathrm{H}} \boldsymbol{n}_k. \quad (2) \end{aligned}$$

The achievable sum rate of this K-user MIMO system under Gaussian signaling is given by

$$\begin{aligned} R_{\mathrm{sum}} = \sum_{k=1}^{K} &\ \log \det(\boldsymbol{I}_{N_{s_k}} + \boldsymbol{G}_{D_k}^{\mathrm{H}} \boldsymbol{G}_{A_k}^{\mathrm{H}} \boldsymbol{H}_{k,k} \boldsymbol{F}_{A_k} \boldsymbol{F}_{D_k} \\ &\cdot (\boldsymbol{G}_{D_k}^{\mathrm{H}} \boldsymbol{G}_{A_k}^{\mathrm{H}} \boldsymbol{H}_{k,k} \boldsymbol{F}_{A_k} \boldsymbol{F}_{D_k})^{\mathrm{H}} \boldsymbol{R}_k^{-1}), \quad (3) \end{aligned}$$

where $\boldsymbol{R}_k = \boldsymbol{G}_{D_k}^{\mathrm{H}} \boldsymbol{G}_{A_k}^{\mathrm{H}} \big(\sum_{i \neq k} \boldsymbol{H}_{k,i} \boldsymbol{F}_{A_i} \boldsymbol{F}_{D_i} \boldsymbol{F}_{D_i}^{\mathrm{H}} \boldsymbol{F}_{A_i}^{\mathrm{H}} \boldsymbol{H}_{k,i}^{\mathrm{H}} + \sigma_{n_k}^2 \boldsymbol{I}_{N_{r_k}} \big) \boldsymbol{G}_{A_k} \boldsymbol{G}_{D_k}$ is the covariance matrix of the inter-user interference plus noise at the kth receiver, $\forall k$. We aim to jointly design hybrid precoders and combiners $\mathcal{A} = \{\mathcal{A}_k, \forall k\}$ with $\mathcal{A}_k = \{\boldsymbol{F}_{A_k}, \boldsymbol{F}_{D_k}, \boldsymbol{G}_{D_k}, \boldsymbol{G}_{A_k}\}$ to maximize the achievable sum rate R_{sum} in (3), which is formulated as

$$\begin{aligned} \max_{\mathcal{A}_k, \forall k} \quad & R_{\mathrm{sum}}, \\ \text{s.t.} \quad & \mathrm{Tr}(\boldsymbol{F}_{A_k} \boldsymbol{F}_{D_k} \boldsymbol{F}_{D_k}^{\mathrm{H}} \boldsymbol{F}_{A_k}^{\mathrm{H}}) \leq P_k, \\ & |[\boldsymbol{F}_{A_k}]_{n,m}|^2 = 1, |[\boldsymbol{G}_{A_k}]_{n,m}|^2 = 1, \ \forall k, n, m. \end{aligned} \quad (4)$$

Clearly, the sum rate maximization problem (4) is nonconvex and NP-hard w.r.t. $\mathcal{A}$ due to the coupled optimization variables and unit-modulus constraints. Moreover, even the optimal fully-digital solution $\boldsymbol{F}_k = \boldsymbol{F}_{A_k} \boldsymbol{F}_{D_k}$ to problem (4) without unit-modulus constraints has not been globally addressed yet, and only stationary solution generated from iterative process is available [40]. Therefore, for K-user MIMO interference systems, the traditional method of minimizing the Euclidean distance between the hybrid analog-digital precoder and the optimal fully-digital one cannot theoretically guarantee its sum rate performance. In the sequel, based on a reasonable assumption on the analog precoder in large-scale MIMO systems, we propose an effective joint design of hybrid precoder and combiner by applying the MM-based alternating optimization with guaranteed sum rate performance to problem (4).

Although the proposed alternating optimization algorithm achieves the semi closed-form solution to each subproblem, it imposes heavy coordination among all transmit-receive pairs. To further reduce complexity, a two-stage hybrid design is firstly proposed with the decoupled optimization of analog and digital precoder/combiner for each transmit-receive pair. In addition, two kinds of hybrid designs using the BD-ZF and SLNR-Max fully-digital precoders are also studied, both of which support the independent hybrid precoder and combiner designs. All the above schemes require global CSI at transmitter, which imposes huge training and feedback overhead. To alleviate this issue, we also consider the partially connected hybrid transceiver structure with significantly reduced feedback overhead and hardware cost, to which the proposed MM-based alternating optimization is directly applicable and a stationary solution of problem (4) can be achieved.

B. Channel Model

In our work, two kinds of block-fading channels are adopted, mmWave channels and Rayleigh fading channels. The first type considers the propagation environment at the mmWave band, which has limited scattering and suffers from several blockage and reduced diffraction, while the other considers the propagation environment with rich scatterers. Moreover, to make the system capacity independent of the scaling of the channel matrix, the normalized channel matrix
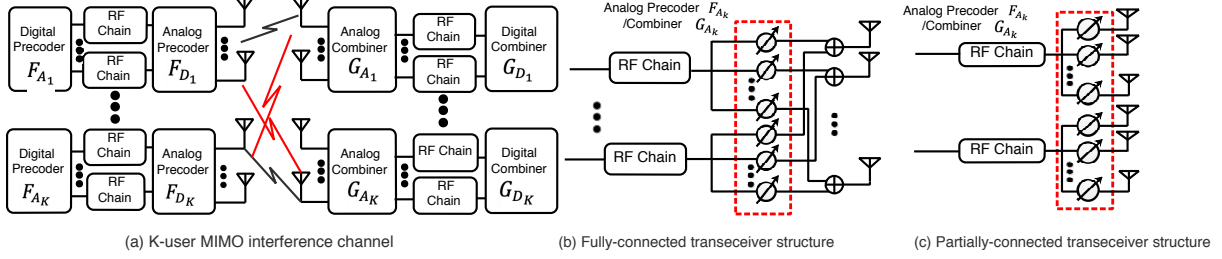
Fig. 1. A K-user MIMO interference channel with different hybrid transceiver structures.

is considered. For Rayleigh fading channels, the elements of the channel matrix $\boldsymbol{H}_{k,i}$ are assumed to be i.i.d. complex Gaussian variables with zero mean and unit variance, i.e., $\text{vec}(\boldsymbol{H}_{k,i}) \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_{N_{r_k} N_{t_i}})$, $\forall k, i$, while for mmWave channels the extended Salen-Valenzuela geometric model [41] is adopted:

$$\boldsymbol{H}_{k,i} = \sqrt{\frac{N_{r_k} N_{t_i}}{L_{k,i}}} \sum_{l=1}^{L_{k,i}} \alpha_k^l \boldsymbol{a}_r(\theta_k^l) \boldsymbol{a}_t^{\text{H}}(\psi_i^l), \ \forall \ k, i, \quad (5)$$

where $L_{k,i}$ denotes the number of dominated propagation paths in the channel $\boldsymbol{H}_{k,i}$ and α_k^l is the complex gain of the lth path, while θ_k^l and ψ_i^l are the angle of arrival (AOA) and angle of departure (AOD) of the lth path, respectively. Assume that the uniform linear array (ULA) with half-wavelength element spacing is deployed at each transmit-receive pair. The transmit and receive array steering vectors can then be expressed as $\boldsymbol{a}_t(\psi_i^l) = \frac{1}{\sqrt{N_{t_i}}} \begin{bmatrix} 1 & e^{-\text{j}\pi \sin \psi_i^l} & \cdots & e^{-\text{j}(N_{t_i}-1)\pi \sin \psi_i^l} \end{bmatrix}^{\text{T}}$ and $\boldsymbol{a}_r(\theta_k^l) = \frac{1}{\sqrt{N_{r_k}}} \begin{bmatrix} 1 & e^{-\text{j}\pi \sin \theta_k^l} & \cdots & e^{-\text{j}(N_{r_k}-1)\pi \sin \theta_k^l} \end{bmatrix}^{\text{T}}$, respectively.

C. Joint Hybrid Transceiver Design

To tackle the sum rate maximization problem (4) effectively, we introduce $\widetilde{\boldsymbol{F}}_{D_k} = (\boldsymbol{F}_{A_k}^{\text{H}} \boldsymbol{F}_{A_k})^{\frac{1}{2}} \boldsymbol{F}_{D_k}$ and $\widetilde{\boldsymbol{F}}_{A_k} = \boldsymbol{F}_{A_k} (\boldsymbol{F}_{A_k}^{\text{H}} \boldsymbol{F}_{A_k})^{-\frac{1}{2}}$, $\forall k$, and refer to [40] to equivalently reformulate it as the following WMMSE problem

$$\min_{\mathcal{A}_k, \boldsymbol{W}_k \succ \boldsymbol{0}} \sum_{k=1}^{K} \left(\text{Tr}(\boldsymbol{W}_k \boldsymbol{E}_k(\mathcal{A}_k)) - \log \det(\boldsymbol{W}_k) - N_{s_k} \right),$$
$$\text{s.t.} \ \ \text{Tr}(\widetilde{\boldsymbol{F}}_{D_k}^{\text{H}} \widetilde{\boldsymbol{F}}_{D_k}) \leq P_k, \ \widetilde{\boldsymbol{F}}_{A_k} = \boldsymbol{F}_{A_k} (\boldsymbol{F}_{A_k}^{\text{H}} \boldsymbol{F}_{A_k})^{-\frac{1}{2}}, \quad (6)$$
$$\left| [\boldsymbol{F}_{A_k}]_{n,m} \right| = 1, \ \left| [\boldsymbol{G}_{A_k}]_{n,m} \right| = 1, \ \forall k, n, m,$$

where the notation $\mathcal{A}_k = \{\boldsymbol{G}_{A_k}, \boldsymbol{G}_{D_k}, \widetilde{\boldsymbol{F}}_{A_k}, \widetilde{\boldsymbol{F}}_{D_k}\}$ is redefined and the MSE matrix $\boldsymbol{E}_k(\mathcal{A}_k)$ is given by

$$\boldsymbol{E}_k(\mathcal{A}_k) = \mathbb{E}\left[(\widehat{\boldsymbol{s}}_k - \boldsymbol{s}_k)(\widehat{\boldsymbol{s}}_k - \boldsymbol{s}_k)^{\text{H}} \right]$$
$$= (\boldsymbol{G}_{D_k}^{\text{H}} \boldsymbol{G}_{A_k}^{\text{H}} \boldsymbol{H}_{k,k} \widetilde{\boldsymbol{F}}_{A_k} \widetilde{\boldsymbol{F}}_{D_k} - \boldsymbol{I}_{N_{s_k}})(\boldsymbol{G}_{D_k}^{\text{H}} \boldsymbol{G}_{A_k}^{\text{H}} \boldsymbol{H}_{k,k} \widetilde{\boldsymbol{F}}_{A_k} \widetilde{\boldsymbol{F}}_{D_k} - \boldsymbol{I}_{N_{s_k}})^{\text{H}}$$
$$+ \sum_{i \neq k} (\boldsymbol{G}_{D_k}^{\text{H}} \boldsymbol{G}_{A_k}^{\text{H}} \boldsymbol{H}_{k,i} \widetilde{\boldsymbol{F}}_{A_i} \widetilde{\boldsymbol{F}}_{D_i})(\boldsymbol{G}_{D_k}^{\text{H}} \boldsymbol{G}_{A_k}^{\text{H}} \boldsymbol{H}_{k,i} \widetilde{\boldsymbol{F}}_{A_i} \widetilde{\boldsymbol{F}}_{D_i})^{\text{H}}$$
$$+ \sigma_{n_k}^2 \boldsymbol{G}_{D_k}^{\text{H}} \boldsymbol{G}_{A_k}^{\text{H}} \boldsymbol{G}_{A_k} \boldsymbol{G}_{D_k}. \quad (7)$$

The proof of equivalence between problem (6) and problem (4) is relegated to Appendix A. Suffice to say that the newly defined variables $\widetilde{\boldsymbol{F}}_{A_k}$ and $\widetilde{\boldsymbol{F}}_{D_k}$ can be applied to tackle coupled nature of the analog precoder $\boldsymbol{F}_{A_k}$ and digital precoder $\boldsymbol{F}_{D_k}$ in the transmit power constraint of problem (4). Based on this, all constraints in the equivalent WMMSE problem (6) become separable. Nevertheless, since the unit-modulus constraints cannot be explicitly reformulated w.r.t $\widetilde{\boldsymbol{F}}_{A_k}$, problem (6) is still challenging to address. To make

problem (6) more tractable, a reasonable approximation of $\widetilde{\boldsymbol{F}}_{A_k}$ is considered. Specifically, it is known from [7], [12], [28] that when $N_{t_k} \to \infty$, the analog precoder design conceived for approximating the near-optimal system performance typically satisfies $\boldsymbol{F}_{A_k}^{\text{H}} \boldsymbol{F}_{A_k} \approx N_{t_k} \boldsymbol{I}_{N_{t_k}^{RF}}$ with high probability, based on which $\widetilde{\boldsymbol{F}}_{A_k}$ can be relaxed to $\widetilde{\boldsymbol{F}}_{A_k} \approx \frac{1}{\sqrt{N_{t_k}}} \boldsymbol{F}_{A_k}$, and problem (6) is then simplified as

$$\min_{\widetilde{\mathcal{A}}_k, \boldsymbol{W}_k \succ \boldsymbol{0}} \sum_{k=1}^{K} \left(\text{Tr}(\boldsymbol{W}_k \boldsymbol{E}_k(\widetilde{\mathcal{A}}_k)) - \log \det(\boldsymbol{W}_k) - N_{s_k} \right),$$
$$\text{s.t.} \ \ \text{Tr}(\widetilde{\boldsymbol{F}}_{D_k}^{\text{H}} \widetilde{\boldsymbol{F}}_{D_k}) \leq P_k, \ \ \left| [\boldsymbol{F}_{A_k}]_{n,m} \right| = 1, \quad (8)$$
$$\left| [\boldsymbol{G}_{A_k}]_{n,m} \right| = 1, \forall k, n, m,$$

where $\widetilde{\mathcal{A}}_k = \{\boldsymbol{G}_{A_k}, \boldsymbol{G}_{D_k}, \boldsymbol{F}_{A_k}, \widetilde{\boldsymbol{F}}_{D_k}\}$ and $\boldsymbol{E}_k(\widetilde{\mathcal{A}}_k)$ is obtained by considering $\widetilde{\boldsymbol{F}}_{A_k} \approx \frac{1}{\sqrt{N_{t_k}}} \boldsymbol{F}_{A_k}$ in (7). We aim for jointly optimizing the hybrid precoder and combiner via solving problem (8). Firstly, it is easily observed that problem (8) is convex w.r.t $\boldsymbol{W}_k$ with given $\widetilde{\mathcal{A}}_k, \forall k$ [40]. By setting the first-order derivative of the objective function in problem (8) w.r.t $\boldsymbol{W}_k$ to zero, we have

$$\boldsymbol{W}_k = \boldsymbol{E}_k^{-1}(\widetilde{\mathcal{A}}_k), \ \forall k. \quad (9)$$

Furthermore, as pointed out in Appendix A, the optimal digital combiner $\boldsymbol{G}_{D_k}$ for simultaneously minimizing the MSEs of all data streams of the kth transmit-receive pair is the Wiener filter:

$$\boldsymbol{G}_{D_k} = \frac{1}{\sqrt{N_{t_k}}} \widetilde{\boldsymbol{Q}}_k \boldsymbol{G}_{A_k}^{\text{H}} \boldsymbol{H}_{k,k} \boldsymbol{F}_{A_k} \widetilde{\boldsymbol{F}}_{D_k}, \ \forall k, \quad (10)$$

where $\widetilde{\boldsymbol{Q}}_k = \left(\sum_{i=1}^{K} \frac{1}{N_{t_i}} (\boldsymbol{G}_{A_k}^{\text{H}} \boldsymbol{H}_{k,i} \boldsymbol{F}_{A_i} \widetilde{\boldsymbol{F}}_{D_i})(\boldsymbol{G}_{A_k}^{\text{H}} \boldsymbol{H}_{k,i} \boldsymbol{F}_{A_i} \widetilde{\boldsymbol{F}}_{D_i})^{\text{H}} + \sigma_{n_k}^2 \boldsymbol{G}_{A_k}^{\text{H}} \boldsymbol{G}_{A_k} \right)^{-1}$. Using the closed-form solutions in (9) and (10), our next task is to find the optimal solution $\{\boldsymbol{F}_{A_k}, \boldsymbol{F}_{D_k}, \boldsymbol{G}_{A_k}, \forall k\}$ to the nonconvex problem (8) via the proposed MM-based alternating optimization algorithm in Section III.

III. Proposed MM-Based Alternating Optimization

A. Brief Review of MM Method

The MM method is an effective optimization tool for solving nonconvex problems. The basic idea is to transform the original nonconvex problem into a sequence of majorized subproblems that can be solved with semi closed-from solutions and guaranteed convergence. The MM method generally consists of two stages, the majorization stage and the minimization stage. In the majorization stage, for a general optimization problem

$$\min_{\boldsymbol{X}} \ f(\boldsymbol{X}), \ \text{s.t.} \ \boldsymbol{X} \in \mathcal{X}, \tag{11}$$

where $\mathcal{X}$ is a closed nonempty set. In terms of our work, it can be nonconvex. Our aim is to find a continuous surrogate function $g(\boldsymbol{X}; \boldsymbol{X}^{(l)})$, also defined as a majorizer of $f(\boldsymbol{X})$ at $\boldsymbol{X}^{(l)}$, for updating $\boldsymbol{X}$ at the lth iteration. Mathematically, this is expressed as

$$\boldsymbol{X}^{(l+1)} = \arg \min_{\boldsymbol{X} \in \mathcal{X}} \ g(\boldsymbol{X}; \boldsymbol{X}^{(l)}). \tag{12}$$

The majorizer $g(\boldsymbol{X}; \boldsymbol{X}^{(l)})$ must satisfy the following conditions to ensure that the MM method converges to a stationary point of problem (11) [35]:

$$\begin{cases} g(\boldsymbol{X}; \boldsymbol{X}^{(l)}) \geq f(\boldsymbol{X}), \ \forall \ \boldsymbol{X} \in \mathcal{X}, \\ g(\boldsymbol{X}^{(l)}; \boldsymbol{X}^{(l)}) = f(\boldsymbol{X}^{(l)}), \ \forall \ \boldsymbol{X}^{(l)} \in \mathcal{X}, \\ g^{'}(\boldsymbol{X}^{(l)}; \boldsymbol{X}^{(l)}|\boldsymbol{d}) = f^{'}(\boldsymbol{X}^{(l)}|\boldsymbol{d}), \ \forall \boldsymbol{d} \in \mathbb{T}_{\mathcal{X}}(\boldsymbol{X}^{(l)}), \end{cases} \tag{13}$$

where $\mathbb{T}_{\mathcal{X}}(\boldsymbol{X}^{(l)})$ is the Boulingand tangent cone [39] of $\mathcal{X}$ at $\boldsymbol{X}^{(l)}$. It is known that the limit point obtained by minimizing $g(\boldsymbol{X}; \boldsymbol{X}^{(l)})$ subject to $\boldsymbol{X} \in \mathcal{X}$ satisfies the stationary condition $f^{'}(\boldsymbol{X}^{(\infty)}|\boldsymbol{d}) \geq 0, \ \forall \boldsymbol{d} \in \mathbb{T}_{\mathcal{X}}(\boldsymbol{X}^{(\infty)})$. Also, based on (13), the monotonicity of the MM method is manifested by

$$f(\boldsymbol{X}^{(l+1)}) \leq g(\boldsymbol{X}^{(l+1)}; \boldsymbol{X}^{(l)}) \leq g(\boldsymbol{X}^{(l)}; \boldsymbol{X}^{(l)}) = f(\boldsymbol{X}^{(l)}), \forall l. \tag{14}$$

The interested readers can refer to [35], [38] for more details on the general MM method.

B. Proposed MM-based Alternating Optimization

In our proposed MM-based alternating optimization for problem (8), we first partition the remaining variables into three blocks, namely $\{\widetilde{\boldsymbol{F}}_{D_k}, \forall k\}$, $\{\boldsymbol{F}_{A_k}, \forall k\}$ and $\{\boldsymbol{G}_{A_k}, \forall k\}$. The MM method is then utilized to update the blocks $\{\boldsymbol{F}_{A_k}, \forall k\}$ and $\{\boldsymbol{G}_{A_k}, \forall k\}$, respectively, while keeping the other blocks fixed. Compared to directly applying the MM method to problem (8) with a single block of all variables, this approach provides more flexibility in designing surrogate functions for better approximating the objective function of problem (8), and thus leads to faster convergence rate [35].

1) Semi Closed-Form Digital Precoder $\{\widetilde{\boldsymbol{F}}_{D_k}, \forall k\}$: Given the fixed $\{\boldsymbol{F}_{A_k}, \boldsymbol{G}_{A_k}, \boldsymbol{G}_{D_k}, \boldsymbol{W}_k, \forall k\}$, by omitting the constant term, the objective function of problem (8) can be rewritten in terms of $\widetilde{\boldsymbol{F}}_{D_k}$ as

$$\begin{aligned} F_{\text{obj}}(\widetilde{\mathcal{A}}) &= \sum_{k=1}^{K} \text{Tr}\big(\boldsymbol{W}_k \boldsymbol{E}_k(\widetilde{\mathcal{A}}_k)\big) \\ &= \sum_{k=1}^{K} \sum_{i=1}^{K} \text{Tr}\big(\widetilde{\boldsymbol{F}}_{D_k}^{\text{H}} \boldsymbol{L}_{i,k}^{\text{H}} \boldsymbol{W}_i \boldsymbol{L}_{i,k} \widetilde{\boldsymbol{F}}_{D_k}\big) \\ &\quad - \sum_{k=1}^{K} \text{Tr}\big(\boldsymbol{W}_k \boldsymbol{L}_{k,k} \widetilde{\boldsymbol{F}}_{D_k} + \boldsymbol{W}_k \widetilde{\boldsymbol{F}}_{D_k}^{\text{H}} \boldsymbol{L}_{k,k}^{\text{H}}\big) + C_1, \end{aligned} \tag{15}$$

where $\boldsymbol{L}_{i,k} = \frac{1}{\sqrt{N_{t_k}}} \boldsymbol{G}_{D_i}^{\text{H}} \boldsymbol{G}_{A_i}^{\text{H}} \boldsymbol{H}_{i,k} \boldsymbol{F}_{A_k}, \ \forall i, k$, and $C_1 = \sum_{k=1}^{K} \text{Tr}(\boldsymbol{W}_k + \sigma_{n_k}^2 \boldsymbol{G}_{A_k} \boldsymbol{G}_{D_k} \boldsymbol{W}_k \boldsymbol{G}_{D_k}^{\text{H}} \boldsymbol{G}_{A_k}^{\text{H}})$. Further, by taking the first-order derivative of $F_{\text{obj}}(\widetilde{\mathcal{A}})$ w.r.t. $\widetilde{\boldsymbol{F}}_{D_k}$, we have

$$\widetilde{\boldsymbol{F}}_{D_k} = \Big(\sum_{i=1}^{K} \boldsymbol{L}_{i,k}^{\text{H}} \boldsymbol{W}_i \boldsymbol{L}_{i,k} + \beta_k \boldsymbol{I}_{N_{t_k}^{RF}}\Big)^{-1} \boldsymbol{L}_{k,k}^{\text{H}} \boldsymbol{W}_k, \ \forall k, \tag{16}$$

where β_k is the dual variable associated with the kth transmit power constraint and its optimal value satisfies complementary slackness condition $\beta_k\big(\text{Tr}(\widetilde{\boldsymbol{F}}_{D_k} \widetilde{\boldsymbol{F}}_{D_k}^{\text{H}}) - P_k\big) = 0$.

To be specific, if $\text{Tr}(\widetilde{\boldsymbol{F}}_{D_k} \widetilde{\boldsymbol{F}}_{D_k}^{\text{H}}) \leq P_k$ holds, we have the optimal $\beta_k^{\text{opt}} = 0$; otherwise, β_k^{opt} can be derived from $\sum_{m=1}^{N_{t_k}^{RF}} \frac{[\boldsymbol{Q}_k \boldsymbol{Q}_k^{\text{H}}]_{m,m}}{([\boldsymbol{\Lambda}_{L_k}]_{m,m} + \beta_k^{\text{opt}})^2} = P_k$, where $\boldsymbol{Q}_k = \boldsymbol{U}_{L_k}^{\text{H}} \boldsymbol{L}_{k,k}^{\text{H}} \boldsymbol{W}_k$ and $\boldsymbol{\Lambda}_{L_k}$ follows from the eigenvalue decomposition (EVD) $\sum_{i=1}^{K} \boldsymbol{L}_{i,k}^{\text{H}} \boldsymbol{W}_i \boldsymbol{L}_{i,k} = \boldsymbol{U}_{L_k} \boldsymbol{\Lambda}_{L_k} \boldsymbol{U}_{L_k}^{\text{H}}$.

2) Semi Closed-Form Analog Precoder $\{\boldsymbol{F}_{A_k}, \forall k\}$: When $\{\widetilde{\boldsymbol{F}}_{D_k}, \boldsymbol{G}_{A_k}, \boldsymbol{G}_{D_k}, \boldsymbol{W}_k, \forall k\}$ are fixed, problem (8) is also non-convex on $\boldsymbol{F}_{A_k}, \forall k$ due to the intractable unit-modulus constraints. To tackle this nonconvexity, we aim to find an effective majorizer of the objective function in problem (8) at $\boldsymbol{F}_{A_k}, \ \forall k$ using the MM method, based on which at least a stationary solution of $\{\boldsymbol{F}_{A_k}, \forall k\}$ can be obtained. According to the identity $\text{Tr}(\boldsymbol{ABCD}) = \text{vec}(\boldsymbol{A}^{\text{T}})^{\text{T}}(\boldsymbol{D}^{\text{T}} \otimes \boldsymbol{B})\text{vec}(\boldsymbol{C})$, the objective function in problem (8) can be re-expressed as

$$G_{\text{obj}}(\widetilde{\mathcal{A}}) = \sum_{k=1}^{K} \boldsymbol{f}_{A_k}^{\text{H}} \widetilde{\boldsymbol{A}}_k \boldsymbol{f}_{A_k} - 2\Re\{\boldsymbol{a}_k^{\text{H}} \boldsymbol{f}_{A_k}\} + C_1, \tag{17}$$

where $\boldsymbol{f}_{A_k} = \text{vec}(\boldsymbol{F}_{A_k})$, $\widetilde{\boldsymbol{A}}_k = \frac{1}{N_{t_k}} \sum_{i=1}^{K} \big((\widetilde{\boldsymbol{F}}_{D_k}^* \widetilde{\boldsymbol{F}}_{D_k}^{\text{T}}) \otimes (\boldsymbol{M}_{i,k} \boldsymbol{W}_i \boldsymbol{M}_{i,k}^{\text{H}})\big)$, $\boldsymbol{M}_{i,k} = \boldsymbol{H}_{i,k}^{\text{H}} \boldsymbol{G}_{A_i} \boldsymbol{G}_{D_i}$, and $\boldsymbol{a}_k^{\text{H}} = \frac{1}{\sqrt{N_{t_k}}} \text{vec}(\boldsymbol{W}_k^{\text{T}})^{\text{T}} \big(\widetilde{\boldsymbol{F}}_{D_k}^{\text{T}} \otimes \boldsymbol{M}_{k,k}^{\text{H}}\big), \ \forall i, k$. It is clearly observed from (17) that there is no coupling among $\boldsymbol{f}_{A_k}, \forall k$ in $G_{\text{obj}}(\widetilde{\mathcal{A}})$, implying that the designs of $\boldsymbol{f}_{A_k}, \forall k$, are independent of each other and can be realized in parallel.

Lemma 1. *[38] For any two Hermitian matrices* $\boldsymbol{Q}, \boldsymbol{Y} \in \mathbb{C}^{N \times N}$ *satisfying* $\boldsymbol{Q} \succeq \boldsymbol{Y}$, *a majorizer of the quadratic function* $\boldsymbol{x}^{\text{H}} \boldsymbol{Y} \boldsymbol{x}$ *at any point* $\boldsymbol{x}_0 \in \mathbb{C}^N$ *is* $\boldsymbol{x}^{\text{H}} \boldsymbol{Q} \boldsymbol{x} + 2\Re(\boldsymbol{x}^{\text{H}}(\boldsymbol{Y} - \boldsymbol{Q})\boldsymbol{x}_0) + \boldsymbol{x}_0^{\text{H}}(\boldsymbol{Q} - \boldsymbol{Y})\boldsymbol{x}_0$.

According to Lemma 1, a majorizer $g_{\text{obj}}(\boldsymbol{f}_{A_k}; \boldsymbol{f}_{A_k}^{(l)})$ of $G_{\text{obj}}(\widetilde{\mathcal{A}})$ at $\boldsymbol{f}_{A_k}^{(l)}$ can be constructed as

$$\begin{aligned} g_{\text{obj}}(\boldsymbol{f}_{A_k}; \boldsymbol{f}_{A_k}^{(l)}) &= \lambda_{\max}(\widetilde{\boldsymbol{A}}_k) \boldsymbol{f}_{A_k}^{\text{H}} \boldsymbol{f}_{A_k} + 2\Re\{\boldsymbol{f}_{A_k}^{\text{H}} \widetilde{\boldsymbol{a}}_k\} \\ &\quad + (\boldsymbol{f}_{A_k}^{(l)})^{\text{H}}(\lambda_{\max}(\widetilde{\boldsymbol{A}}_k) \boldsymbol{I}_{N_{t_k} N_{t_k}^{RF}} - \widetilde{\boldsymbol{A}}_k) \boldsymbol{f}_{A_k}^{(l)}, \end{aligned} \tag{18}$$

where $\widetilde{\boldsymbol{a}}_k = \big(\widetilde{\boldsymbol{A}}_k - \lambda_{\max}(\widetilde{\boldsymbol{A}}_k) \boldsymbol{I}_{N_{t_k} N_{t_k}^{RF}}\big) \boldsymbol{f}_{A_k}^{(l)} - \boldsymbol{a}_k$. Hence, the majorized problem for optimizing $\boldsymbol{f}_{A_k}$ can be formulated as

$$\min_{\boldsymbol{f}_{A_k}} \ \Re\{\boldsymbol{f}_{A_k}^{\text{H}} \widetilde{\boldsymbol{a}}_k\}, \ \text{s.t.} \ |[\boldsymbol{f}_{A_k}]_q| = 1, \forall q = 1 \cdots N_{t_k} N_{t_k}^{RF}, \tag{19}$$

to which the semi closed-form solution is given by

$$\boldsymbol{f}_{A_k} = -e^{\text{j} \arg(\widetilde{\boldsymbol{a}}_k)}, \ \forall k. \tag{20}$$

3) Semi Closed-Form Analog Combiner $\{\boldsymbol{G}_{A_k}, \forall k\}$: By fixing $\{\boldsymbol{G}_{D_k}, \boldsymbol{F}_{A_k}, \widetilde{\boldsymbol{F}}_{D_k}, \boldsymbol{W}_k, \forall k\}$, we can re-express the objective function of problem (8) w.r.t. $\boldsymbol{G}_{A_k}, \forall k$, as

$$S_{\text{obj}}(\widetilde{\mathcal{A}}) = \sum_{k=1}^{K} \boldsymbol{g}_{A_k}^{\text{H}} \widetilde{\boldsymbol{N}}_k \boldsymbol{g}_{A_k} - 2\Re\{\boldsymbol{d}_k^{\text{H}} \boldsymbol{g}_{A_k}\}, \tag{21}$$

where $\boldsymbol{g}_{A_k} = \text{vec}(\boldsymbol{G}_{A_k})$, $\boldsymbol{d}_k^{\text{H}} = \text{vec}(\boldsymbol{D}_k^{\text{T}} \boldsymbol{G}_{D_k}^{\text{T}})^{\text{T}}$, $\boldsymbol{D}_k = \frac{1}{\sqrt{N_{t_k}}} \boldsymbol{W}_k \widetilde{\boldsymbol{F}}_{D_k}^{\text{H}} \boldsymbol{F}_{A_k}^{\text{H}} \boldsymbol{H}_{k,k}^{\text{H}}$ and $\widetilde{\boldsymbol{N}}_k = (\boldsymbol{G}_{D_k} \boldsymbol{W}_k \boldsymbol{G}_{D_k}^{\text{H}})^{\text{T}} \otimes \big(\sum_{i=1}^{K} \frac{1}{N_{t_k}} \boldsymbol{H}_{k,i} \boldsymbol{F}_{A_i} \widetilde{\boldsymbol{F}}_{D_i} \widetilde{\boldsymbol{F}}_{D_i}^{\text{H}} \boldsymbol{F}_{A_i}^{\text{H}} \boldsymbol{H}_{k,i}^{\text{H}} + \sigma_{n_k}^2 \boldsymbol{I}_{N_{r_k}}\big)$. Obviously, the function $S_{\text{obj}}(\widetilde{\mathcal{A}})$ is separable in terms of $\boldsymbol{g}_{A_k}, \forall k$. Similarly, based on Lemma 1, a majorizer $s_{\text{obj}}(\boldsymbol{g}_{A_k}; \boldsymbol{g}_{A_k}^{(l)})$ of $S_{\text{obj}}(\widetilde{\mathcal{A}})$ at $\boldsymbol{g}_{A_k}^{(l)}$ is given by

$$s_{\text{obj}}(\boldsymbol{g}_{A_k}; \boldsymbol{g}_{A_k}^{(l)}) = \lambda_{\max}(\widetilde{\boldsymbol{N}}_k) \boldsymbol{g}_{A_k}^{\text{H}} \boldsymbol{g}_{A_k} + 2\Re\{\boldsymbol{g}_{A_k}^{\text{H}} \widetilde{\boldsymbol{d}}_k\} + C_3, \tag{22}$$

Algorithm 1 MM-Alt-Opt: Joint hybrid transceiver design for the problem (8)

Input: Initial hybrid precoders and combiners $\widetilde{\mathcal{A}}^{(0)}$; outer iteration index $I_W = 0$; convergence threshold ϵ_{obj}.

1: **repeat**
2:     Calculate $\boldsymbol{G}_{D_k}^{(I_W+1)}, \forall k$, according to (10).
3:     Fix $\boldsymbol{G}_{D_k}^{(I_W+1)}, \forall k$, calculate $\boldsymbol{W}_k^{(I_W+1)}, \forall k$, according to (9).
4:     Fix $\{\boldsymbol{W}_k^{(I_W+1)}, \boldsymbol{G}_{D_k}^{(I_W+1)}, \forall k\}$, calculate $\widetilde{\boldsymbol{F}}_{D_k}^{(I_W+1)}, \forall k$, according to (16).
5:     Update $\widetilde{\mathcal{A}}^{(I_W)}$ with $\{\widetilde{\boldsymbol{F}}_{D_k}^{(I_W+1)}, \boldsymbol{W}_k^{(I_W+1)}, \boldsymbol{G}_{D_k}^{(I_W+1)}, \forall k\}$.
6:     Calculate $\boldsymbol{F}_{A_k}^{(I_W+1)}, \forall k$, using MM method
    $\begin{cases} \textbf{Input } \text{inner iteration index } I_M = 0 \text{ and set } \widetilde{\mathcal{A}}^{(I_M)} = \widetilde{\mathcal{A}}^{(I_W)}. \\ \quad \textbf{while } |F_{\text{obj}}(\widetilde{\mathcal{A}}^{(I_M)}) - F_{\text{obj}}(\widetilde{\mathcal{A}}^{(I_M-1)})| \leq \epsilon_{obj} \textbf{ do} \\ \qquad \text{Calculate } \boldsymbol{F}_{A_k}^{(I_M+1)}, \forall k, \text{ according to (20).} \\ \qquad \text{Set } I_M = I_M + 1 \text{ and update } \widetilde{\mathcal{A}}^{(I_M)} \text{ with } \boldsymbol{F}_{A_k}^{(I_M)}, \forall k. \\ \quad \textbf{end while} \\ \textbf{Output } \boldsymbol{F}_{A_k}^{(I_W+1)} = \boldsymbol{F}_{A_k}^{(I_M)}, \forall k. \end{cases}$
7:     Update $\widetilde{\mathcal{A}}^{(I_W)}$ with $\{\boldsymbol{W}_k^{(I_W+1)}, \boldsymbol{G}_{D_k}^{(I_W+1)}, \widetilde{\boldsymbol{F}}_{D_k}^{(I_W+1)}, \boldsymbol{F}_{A_k}^{(I_W+1)}, \forall k\}$.
8:     Calculate $\boldsymbol{G}_{A_k}^{(I_W+1)}, \forall k$, using MM method
    $\begin{cases} \textbf{Input } \text{inner iteration index } I_M = 0 \text{ and set } \widetilde{\mathcal{A}}^{(I_M)} = \widetilde{\mathcal{A}}^{(I_W)}. \\ \quad \textbf{while } |F_{\text{obj}}(\widetilde{\mathcal{A}}^{(I_M)}) - F_{\text{obj}}(\widetilde{\mathcal{A}}^{(I_M-1)})| \leq \epsilon_{obj} \textbf{ do} \\ \qquad \text{Calculate } \boldsymbol{G}_{A_k}^{(I_M+1)} \text{ according to (24).} \\ \qquad \text{Set } I_M = I_M + 1 \text{ and update } \widetilde{\mathcal{A}}^{(I_M)} \text{ with } \boldsymbol{G}_{A_k}^{(I_M)}, \forall k. \\ \quad \textbf{end while} \\ \textbf{Output } \boldsymbol{G}_{A_k}^{(I_W+1)} = \boldsymbol{G}_{A_k}^{(I_M)}, \forall k. \end{cases}$
9:     Set $\widetilde{\mathcal{A}}^{(I_W+1)} = \{\boldsymbol{W}_k^{(I_W+1)}, \boldsymbol{G}_{D_k}^{(I_W+1)}, \widetilde{\boldsymbol{F}}_{D_k}^{(I_W+1)}, \boldsymbol{F}_{A_k}^{(I_W+1)}, \boldsymbol{G}_{A_k}^{(I_W+1)}, \forall k\}$ and $I_W = I_W + 1$.
10: **until** $|F_{\text{obj}}(\widetilde{\mathcal{A}}^{(I_M)}) - F_{\text{obj}}(\widetilde{\mathcal{A}}^{(I_M-1)})| \leq \epsilon_{obj}$.

Output: $\{\boldsymbol{G}_{D_k}^{(I_W)}, \boldsymbol{G}_{A_k}^{(I_W)}, \boldsymbol{F}_{D_k}^{(I_W)}, \boldsymbol{F}_{A_k}^{(I_W)}, \forall k\}$ based on $\widetilde{\boldsymbol{F}}_{D_k} = \left(\boldsymbol{F}_{A_k}^{\text{H}} \boldsymbol{F}_{A_k}\right)^{\frac{1}{2}} \boldsymbol{F}_{D_k}$.

---

where $\widetilde{\boldsymbol{d}}_k = (\widetilde{\boldsymbol{N}}_k - \lambda_{\max}(\widetilde{\boldsymbol{N}}_k) \boldsymbol{I}_{N_{r_k} N_{r_k}^{RF}}) \boldsymbol{g}_{A_k}^{(l)} - \boldsymbol{d}_k$ and $C_3 = \text{Tr}(\boldsymbol{W}_k) + (\boldsymbol{g}_{A_k}^{(l)})^{\text{H}} (\lambda_{\max}(\boldsymbol{N}_k) \boldsymbol{I}_{N_{r_k} N_{r_k}^{RF}} - \widetilde{\boldsymbol{N}}_k) \boldsymbol{g}_{A_k}^{(l)}$. Hence, the majorized problem for optimizing $\boldsymbol{g}_{A_k}$ can be simplified as

$$\min_{\boldsymbol{g}_{A_k}} \Re\{\boldsymbol{g}_{A_k}^{\text{H}} \widetilde{\boldsymbol{d}}_k\}, \text{s.t. } |[\boldsymbol{g}_{A_k}]_{q'}| = 1, \ \forall q' = 1 \cdots N_{r_k} N_{r_k}^{RF}, \quad (23)$$

with the semi closed-form solution

$$\boldsymbol{g}_{A_k} = - e^{\text{j} \arg\left(\widetilde{\boldsymbol{d}}_k\right)}. \quad (24)$$

Overall, integrating the solutions in (9), (10), (16), (20) and (24) can lead to the proposed MM-based alternating optimization for the hybrid transceiver design in MIMO interference systems, which is summarized in Algorithm 1.

C. Two-Stage Hybrid Transceiver Design

In this subsection, we propose a two-stage hybrid transceiver design with the decoupled analog and digital precoder/combiner optimization. First, we present a useful property in large-scale MIMO interference systems.

Proposition 1. *For large-scale MIMO interference systems operating over rich scattering Rayleigh fading or mmWave channels, the correlation matrices between different channels* $\boldsymbol{H}_{k,k}$ *and* $\boldsymbol{H}_{i,k}, \forall i \neq k$, *satisfy* $\lim_{N_{t_k} \to +\infty} \frac{1}{N_{t_k}} \boldsymbol{H}_{i,k} \boldsymbol{H}_{k,k}^{\text{H}} = \boldsymbol{0}_{N_{r_i} \times N_{r_k}}, \forall i, k, \ i \neq k$ *[28]. Based on this, by defining the singular value decomposition (SVD):* $\boldsymbol{H}_{i,k} = \boldsymbol{U}_{i,k} \boldsymbol{\Lambda}_{i,k} \boldsymbol{V}_{i,k}^H, \forall i, k$,

we further infer that the first $rank(\boldsymbol{H}_{k,k})$ *columns of* $\boldsymbol{V}_{k,k}$ *and the first* $rank(\boldsymbol{H}_{i,k})$ *columns of* $\boldsymbol{V}_{i,k}, \forall i \neq k$, *are asymptotically orthogonal, i.e.,*

$$\lim_{N_{t_k} \to +\infty} \boldsymbol{V}_{i,k}^{\text{H}}(1 : rank(\boldsymbol{H}_{i,k}), :) \boldsymbol{V}_{k,k}(:, 1 : rank(\boldsymbol{H}_{k,k}))$$
$$= \boldsymbol{0}_{rank(\boldsymbol{H}_{i,k}) \times rank(\boldsymbol{H}_{k,k})}, \quad \forall i \neq k. \quad (25)$$

Proof. See Appendix B. □

According to Proposition 1, the right singular subspace of each desired transmit-receive channel is asymptotically orthogonal to those of the corresponding interference channels, which implies that the inter-user interference can be eliminated by large arrays without any loss of MIMO transceiver design freedom for maximizing the data rate of each transmit-receive pair. Inspired by the above fact, in the first stage, we consider independently designing the analog precoder and combiner $\{\boldsymbol{F}_{A_k}, \boldsymbol{G}_{A_k}\}$ of each transmit-receive pair to maximize the effective channel gain. Mathematically, it is formulated as

$$\min_{\boldsymbol{G}_{A_k}} \quad \|\boldsymbol{G}_{A_k}^{\text{H}} \boldsymbol{H}_{k,k} \boldsymbol{F}_{A_k}\|_F^2,$$
$$\text{s.t.} \quad |[\boldsymbol{F}_{A_k}]_{n,m}| = 1, |[\boldsymbol{G}_{A_k}]_{n,m}| = 1, \forall n, m. \quad (26)$$

Notice that the unconstrained optimal solution to problem (26) is easily derived as [28], [31]

$$\boldsymbol{F}_{A_k}^{\text{Unc}} = \boldsymbol{V}_{k,k}(:, 1 : N_{t_k}^{RF}), \ \boldsymbol{G}_{A_k}^{\text{Unc}} = \boldsymbol{U}_{k,k}(:, 1 : N_{r_k}^{RF}). \quad (27)$$

Next, our goal is to design the unit-modulus analog precoder and combiner for sufficiently closely approximating the closed-form solutions in (27) with the capability of achieving the zero inter-user interference when $N_{t_k} \to +\infty$. Mathematically, the unit-modulus analog precoder $\boldsymbol{F}_{A_k}$ is designed such that

$$\min_{\boldsymbol{F}_{A_k}} \|\boldsymbol{F}_{A_k} - \boldsymbol{F}_{A_k}^{\text{Unc}}\|_F^2, \text{ s.t. } |[\boldsymbol{F}_{A_k}]_{n,m}| = 1, \forall n, m. \quad (28)$$

Likewise, the unit-modulus analog combiner $\boldsymbol{G}_{A_k}$ can also be obtained by solving

$$\min_{\boldsymbol{G}_{A_k}} \|\boldsymbol{G}_{A_k} - \boldsymbol{G}_{A_k}^{\text{Unc}}\|_F^2, \text{ s.t. } |[\boldsymbol{G}_{A_k}]_{n,m}| = 1, \forall n, m. \quad (29)$$

Both the two problems can be globally solved by PP [23], which yields the following closed-form solutions

$$\boldsymbol{F}_{A_k}^{\text{PP}} = e^{\text{j} \arg\left(\boldsymbol{F}_{A_k}^{\text{Unc}}\right)}, \ \boldsymbol{G}_{A_k}^{\text{PP}} = e^{\text{j} \arg\left(\boldsymbol{G}_{A_k}^{\text{Unc}}\right)}. \quad (30)$$

In the second digital stage, to further suppress the inter-user interference at all transmit-receive pairs, the WMMSE-based joint optimization of the digital precoder and combiner is studied. Specifically, based on the obtained analog precoder $\boldsymbol{F}_{A_k}$ and combiner $\boldsymbol{G}_{A_k}$ in the first stage, a low-dimensional alternating optimization between the digital precoder $\widetilde{\boldsymbol{F}}_{D_k}$ in (16) and the digital combiner $\boldsymbol{G}_{D_k}$ in (10) is performed, which clearly has lower computational complexity than the proposed MM-based alternating optimization in Section III-B.

This two-stage hybrid design can be regarded as a special case of the MM-based alternating optimization by predetermining the analog precoder and combiner of each transmit-receiver pair as that in (30), and thus only an iterative procedure between the digital precoder $\boldsymbol{F}_{D_k}$ and the digital combiner $\boldsymbol{G}_{D_k}$ is performed. Moreover, considering that the

performance of the proposed MM-based alternating optimization generally depends on the initial point [40], we heuristically choose the analog precoder and combiner design of (30) as a initial point due to its potential in harvesting large array gain. The superior sum rate performance of this two-stage hybrid design will be illustrated by the numerical simulations in Section VI.

IV. Low-Complexity and Low-Cost Hybrid Transceiver Designs

Although the semi closed-form solutions to the hybrid transceiver design can be obtained using the above two alternating optimization algorithms, they generally require extensive coordination among all transmit-receive pairs. In this section, we investigate low-complexity hybrid transceiver designs from the perspectives of decoupling hybrid precoder and combiner designs for each transmit-receive pair and reducing hardware cost, respectively. Notice that in this section the above approximation $\boldsymbol{F}_{A_k}^H \boldsymbol{F}_{A_k} \approx N_{t_k} \boldsymbol{I}_{N_{t_k}^{RF}}$ is not applied.

A. BD-ZF Hybrid Transceiver Design

It is well-known that the BD-ZF precoding is a near-optimal scheme in multiuser massive MIMO systems via completely eliminating the inter-user interference. By assuming $N_{t_k} > \sum_{i=1}^{K} \mathrm{rank}(\boldsymbol{H}_{i,k})$, we can extend it into our studied MIMO interference channels under hybrid implementation. Specifically, by defining the leakage channel for the kth transmit-receive pair as $\widetilde{\boldsymbol{H}}_k = \begin{bmatrix} \boldsymbol{H}_{1,k}^H \cdots \boldsymbol{H}_{k-1,k}^H \boldsymbol{H}_{k+1,k}^H \cdots \boldsymbol{H}_{K,k}^H \end{bmatrix}, \forall k$, an orthonormal basis for the orthogonal complement of $\widetilde{\boldsymbol{H}}_k$ is given by $\widetilde{\boldsymbol{H}}_k^{\perp} \in \mathbb{C}^{N_{t_k} \times L_k}$ with $L_k = (N_{t_k} - \mathrm{rank}(\widetilde{\boldsymbol{H}}_k)) \geq N_{s_k}$ and $(\widetilde{\boldsymbol{H}}_k^{\perp})^H \widetilde{\boldsymbol{H}}_k^{\perp} = \boldsymbol{I}_{L_k}$. Then the fully-digital BD-ZF precoder $\boldsymbol{F}_k^{\mathrm{ZF}}$ at the kth transmitter for eliminating both inter-user and intra-data interference can be expressed as

$$\boldsymbol{F}_k^{\mathrm{ZF}} = \widetilde{\boldsymbol{H}}_k^{\perp} \widetilde{\boldsymbol{V}}_k(:, 1:N_{s_k}) \sqrt{\boldsymbol{\Lambda}_k}, \ \forall k, \tag{31}$$

where $\widetilde{\boldsymbol{V}}_k \in \mathbb{C}^{L_k \times L_k}$ originates from the SVD $\boldsymbol{H}_{k,k} \widetilde{\boldsymbol{H}}_k^{\perp} = \widetilde{\boldsymbol{U}}_k \widetilde{\boldsymbol{\Lambda}}_k \widetilde{\boldsymbol{V}}_k^H$ with $\widetilde{\boldsymbol{\Lambda}}_k = \mathrm{diag}[\widetilde{\lambda}_{k,1}^2, \cdots, \widetilde{\lambda}_{k,L_k}^2]$, and $\boldsymbol{\Lambda}_k = \mathrm{diag}[f_{k,1}, \cdots, f_{k,N_{s_k}}]$ denotes the optimal solution to the following sum rate maximization problem

$$\begin{aligned} \max_{\{f_{k,s}, \forall k,s\}} \quad & \sum_{k=1}^{K} \sum_{s=1}^{N_{s_k}} \log\left(1 + \sigma_{n_k}^{-2} \widetilde{\lambda}_{k,s}^2 f_{k,s}\right), \\ \text{s.t.} \quad & \sum_{s=1}^{N_{s_k}} f_{k,s} \leq P_k, \ \forall k. \end{aligned} \tag{32}$$

It is clear that the optimal solution to problem (32) has a water-filling structure, i.e., $f_{k,s} = \left[\frac{1}{\mu \ln 2} - \frac{\sigma_{n_k}^2}{\widetilde{\lambda}_{k,s}^2}\right]^+, \forall k, l$, where μ is chosen to satisfy $\sum_{l=1}^{N_{s_k}} f_{k,l} = P_k, \forall k$.

1) Iterative-PP Hybrid Precoder Design: In this subsection, we aim to minimize the Euclidean distance between the hybrid precoder of each transmitter and the corresponding fully-digital BD-ZF precoder in (31), which is formulated as

$$\begin{aligned} \min_{\boldsymbol{F}_{A_k}, \boldsymbol{F}_{D_k}} \quad & \|\boldsymbol{F}_k^{\mathrm{ZF}} - \boldsymbol{F}_{A_k} \boldsymbol{F}_{D_k}\|_F^2, \\ \text{s.t.} \quad & |[\boldsymbol{F}_{A_k}]_{n,m}| = 1, \|\boldsymbol{F}_{A_k} \boldsymbol{F}_{D_k}\|_F^2 = P_k, \ \forall n, m, k, \end{aligned} \tag{33}$$

where the maximum transmission power is considered. Using the defined variables $\widetilde{\boldsymbol{F}}_{D_k}$ and $\widetilde{\boldsymbol{F}}_{A_k}, \forall k$ in Section II-C, problem (33) can be rewritten as

$$\begin{aligned} \max_{\widetilde{\boldsymbol{F}}_{A_k}, \widetilde{\boldsymbol{F}}_{D_k}} \quad & \Re\left\{\mathrm{Tr}\left(\widetilde{\boldsymbol{F}}_{A_k} \widetilde{\boldsymbol{F}}_{D_k} (\boldsymbol{F}_k^{\mathrm{ZF}})^H\right)\right\}, \\ \text{s.t.} \quad & |[\boldsymbol{F}_{A_k}]_{n,m}| = 1, \|\widetilde{\boldsymbol{F}}_{D_k}\|_F^2 = P_k, \ \forall n, m, k. \end{aligned} \tag{34}$$

Although problem (34) is much simplified compared to problem (33), it is still challenging to directly design the analog precoder $\boldsymbol{F}_{A_k}$ in the unit-modulus space. We resort to an iterative-PP based method with two key ingredients: unconstrained optimal analog precoder and alternating minimization. The unconstrained optimal analog precoder $\boldsymbol{F}_{A_k}^{\mathrm{Unc}}$ to problem (34) using majorization theory [31] is summarized in the following proposition.

Proposition 2. *The unconstrained optimal analog precoder $\boldsymbol{F}_{A_k}^{\mathrm{Unc}}$ to problem* (34) *is*

$$\boldsymbol{F}_{A_k}^{\mathrm{Unc}} = \boldsymbol{U}_k^{\mathrm{ZF}} \boldsymbol{\Lambda}_{\boldsymbol{F}_{A_k}} \boldsymbol{V}_{\boldsymbol{F}_{A_k}}, \tag{35}$$

where the unitary matrix $\boldsymbol{U}_k^{\mathrm{ZF}}$ comes from the SVD $\boldsymbol{F}_k^{\mathrm{ZF}} = \boldsymbol{U}_k^{\mathrm{ZF}} \boldsymbol{\Lambda}_k^{\mathrm{ZF}} \boldsymbol{V}_k^{\mathrm{ZF}}$. Moreover, both the diagonal matrix $\boldsymbol{\Lambda}_{\boldsymbol{F}_{A_k}}$ and the unitary matrix $\boldsymbol{V}_{\boldsymbol{F}_{A_k}}$ can be arbitrarily chosen.

Proof. Define the SVDs $\boldsymbol{F}_{A_k} = \boldsymbol{U}_{\boldsymbol{F}_{A_k}} \boldsymbol{\Lambda}_{\boldsymbol{F}_{A_k}} \boldsymbol{V}_{\boldsymbol{F}_{A_k}}$ and $\widetilde{\boldsymbol{F}}_{D_k} = \boldsymbol{U}_{\widetilde{\boldsymbol{F}}_{D_k}} \boldsymbol{\Lambda}_{\widetilde{\boldsymbol{F}}_{D_k}} \boldsymbol{V}_{\widetilde{\boldsymbol{F}}_{D_k}}$, where $\{\boldsymbol{U}_{\boldsymbol{F}_{A_k}}, \boldsymbol{U}_{\widetilde{\boldsymbol{F}}_{D_k}}\}$ and $\{\boldsymbol{V}_{\boldsymbol{F}_{A_k}}, \boldsymbol{V}_{\widetilde{\boldsymbol{F}}_{D_k}}\}$ are the sets of unitary matrices, while $\{\boldsymbol{\Lambda}_{\boldsymbol{F}_{A_k}}, \boldsymbol{\Lambda}_{\widetilde{\boldsymbol{F}}_{D_k}}\}$ are the corresponding diagonal matrices with diagonal elements arranged in a decreasing order. It is observed from (34) that $\widetilde{\boldsymbol{F}}_{D_k}$ subject to the power constraint $\|\widetilde{\boldsymbol{F}}_{D_k}\|_F^2 = P_k$ is unitarily invariant. In other words, both the unitary matrices $\boldsymbol{U}_{\widetilde{\boldsymbol{F}}_{D_k}}$ and $\boldsymbol{V}_{\widetilde{\boldsymbol{F}}_{D_k}}$ are unconstrained. In addition, observing from $\widetilde{\boldsymbol{F}}_{A_k} = \boldsymbol{F}_{A_k} (\boldsymbol{F}_{A_k}^H \boldsymbol{F}_{A_k})^{-\frac{1}{2}} = \boldsymbol{U}_{\boldsymbol{F}_{A_k}} \boldsymbol{V}_{\boldsymbol{F}_{A_k}}$, we find that the diagonal matrix $\boldsymbol{\Lambda}_{\boldsymbol{F}_{A_k}}$ actually has no effect on the maximum value of the objective function in problem (34). Further based on [42, B.2. Theorem (Fan,1951)], the unconstrained optimal analog precoder $\boldsymbol{F}_{A_k}^{\mathrm{Unc}}$ to problem (34) is readily derived as that in (35), where $\boldsymbol{\Lambda}_{\boldsymbol{F}_{A_k}}$ and $\boldsymbol{V}_{\boldsymbol{F}_{A_k}}$ can be arbitrarily chosen. $\square$

Based on Proposition 2, we then aim to find an unit-modulus analog precoder $\boldsymbol{F}_{A_k}$ with the minimum Euclidean distance to the unconstrained optimal $\boldsymbol{F}_{A_k}^{\mathrm{Unc}}$, which is formulated as

$$\begin{aligned} \min_{\boldsymbol{\Lambda}_{\boldsymbol{F}_{A_k}}, \boldsymbol{V}_{\boldsymbol{F}_{A_k}}, \boldsymbol{F}_{A_k}} \quad & \|\boldsymbol{F}_{A_k}^{\mathrm{Unc}} - \boldsymbol{F}_{A_k}\|_F^2 = \|\boldsymbol{U}_k^{\mathrm{ZF}} \boldsymbol{\Lambda}_{\boldsymbol{F}_{A_k}} \boldsymbol{V}_{\boldsymbol{F}_{A_k}} - \boldsymbol{F}_{A_k}\|_F^2, \\ \text{s.t.} \quad & |[\boldsymbol{F}_{A_k}]_{n,m}| = 1, \quad \forall n, m, k. \end{aligned} \tag{36}$$

Since the diagonal matrix $\boldsymbol{\Lambda}_{\boldsymbol{F}_{A_k}}$ has no effect on problem (33), hereafter, we consider the unconstrained diagonal matrix $\boldsymbol{\Lambda}_{\boldsymbol{F}_{A_k}}$. Although problem (36) is not jointly convex w.r.t $\{\boldsymbol{\Lambda}_{\boldsymbol{F}_{A_k}}, \boldsymbol{V}_{\boldsymbol{F}_{A_k}}, \boldsymbol{F}_{A_k}\}$, it is a 'semi-convex' problem, in which the closed-form solution of each variable is easily obtained when fixing all the others. Specifically, given $\boldsymbol{\Lambda}_{\boldsymbol{F}_{A_k}}$ and $\boldsymbol{V}_{\boldsymbol{F}_{A_k}}$, the optimal analog precoder $\boldsymbol{F}_{A_k}$ to problem (36) is readily derived as

$$\boldsymbol{F}_{A_k} = e^{\mathrm{j}\arg\left(\boldsymbol{U}_k^{\mathrm{ZF}} \boldsymbol{\Lambda}_{\boldsymbol{F}_{A_k}} \boldsymbol{V}_{\boldsymbol{F}_{A_k}}\right)}, \ \forall k. \tag{37}$$

By fixing $\boldsymbol{F}_{A_k}$ and $\boldsymbol{V}_{\boldsymbol{F}_{A_k}}$, the optimal diagonal matrix $\boldsymbol{\Lambda}_{\boldsymbol{F}_{A_k}}$ to problem (36) can be obtained by solving the following problem

$$\max_{\boldsymbol{\Lambda}_{\boldsymbol{F}_{A_k}}} \quad \Re\big\{\mathrm{Tr}\big(\boldsymbol{V}_{\boldsymbol{F}_{A_k}}\boldsymbol{F}_{A_k}^{\mathrm{H}}\boldsymbol{U}_k^{\mathrm{ZF}}\boldsymbol{\Lambda}_{\boldsymbol{F}_{A_k}}\big)\big\},$$

$$\text{s.t.} \quad |[\boldsymbol{F}_{A_k}]_{n,m}| = 1, \forall n, m, k, \tag{38}$$

whose closed-form solution is given by

$$\big[\boldsymbol{\Lambda}_{\boldsymbol{F}_{A_k}}\big]_{i,i} = \Re\big\{\big[\boldsymbol{V}_{\boldsymbol{F}_{A_k}}\boldsymbol{F}_{A_k}^{\mathrm{H}}\boldsymbol{U}_k^{\mathrm{ZF}}\big]_{i,i}\big\}, \ i=1,\cdots,N_{t_k}^{RF}. \tag{39}$$

Finally, for the given $\boldsymbol{\Lambda}_{\boldsymbol{F}_{A_k}}$ and $\boldsymbol{F}_{A_k}$, the optimal unitary matrix $\boldsymbol{V}_{\boldsymbol{F}_{A_k}}$ is easily

$$\boldsymbol{V}_{\boldsymbol{F}_{A_k}} = \boldsymbol{V}_{A_k}^{\mathrm{H}}\boldsymbol{U}_{A_k}^{\mathrm{H}}, \ \forall k, \tag{40}$$

where the unitary matrices $\boldsymbol{V}_{A_k}$ and $\boldsymbol{U}_{A_k}$ come from the SVD $\boldsymbol{F}_{A_k}^{\mathrm{H}}\boldsymbol{U}_k^{\mathrm{ZF}}\boldsymbol{\Lambda}_{\boldsymbol{F}_{A_k}} = \boldsymbol{U}_{A_k}\boldsymbol{\Lambda}_{A_k}\boldsymbol{V}_{A_k}$.

Through an alternating optimization among (37), (39) and (40), the iterative PP-based unit-modulus analog precoder $\boldsymbol{F}_{A_k}$ can be finally obtained. Then by applying Lagrangian multiplier method to problem (33), the optimal digital precoder $\boldsymbol{F}_{D_k}$ given $\boldsymbol{F}_{A_k}$ is expressed as

$$\boldsymbol{F}_{D_k} = \frac{\sqrt{P_k}\boldsymbol{F}_{A_k}^{H}\boldsymbol{F}_k^{\mathrm{ZF}}}{\|(\boldsymbol{F}_{A_k}^{H}\boldsymbol{F}_{A_k})^{\frac{1}{2}}\boldsymbol{F}_{A_k}^{H}\boldsymbol{F}_k^{\mathrm{ZF}}\|_F}, \ \forall k. \tag{41}$$

Even when the distance between the hybrid precoder and the fully-digital BD-ZF precoder is minimized, we still cannot guarantee the hybrid precoder's capability of realizing zero inter-user interference, since it may not be exactly located in the null-space of the corresponding leakage channels. However, the effectiveness of the above iterative-PP hybrid precoder design on suppressing the inter-user interference is demonstrated in the following proposition.

Proposition 3. *For mmWave channels, once the hybrid precoder $\boldsymbol{F}_{A_k}\boldsymbol{F}_{D_k}$ is obtained from* (37) *and* (41)*, the resultant inter-user interference to the ith receiver, where $i \neq k$, satisfies*

$$\lim_{N_{t_k}\to+\infty} \boldsymbol{H}_{i,k}\boldsymbol{F}_{A_k}\boldsymbol{F}_{D_k} = \boldsymbol{0}, \ \forall k=1,\cdots,K, \ i\neq k. \tag{42}$$

Proof. See Appendix C. $\qquad\square$

Proposition 3 reveals that in large-scale mmWave scenarios, the above iterative-PP hybrid precoder also achieves the near-zero inter-user interference, like the fully-digital BD-ZF precoder. However, it may not work well in Rayleigh channels due to the following two reasons. The fully-digital BD-ZF precoder with a strict restriction on the numbers of transmit and receive antennas may be infeasible in rich scattering scenarios, especially for a large number of transmit-receive pairs. Also Proposition 3 is not applicable to the channels without sparsity. These facts motivate us to propose another more general low-complexity hybrid precoder design in Section IV-B.

2) MM-based Hybrid Combiner Design: Given the hybrid precoder (37) and (41), the MMSE hybrid combiner design is then formulated as

$$\min_{\boldsymbol{G}_{A_k},\boldsymbol{G}_{D_k}} \quad \mathbb{E}\big[\|\boldsymbol{s}_k - \boldsymbol{G}_{D_k}^{\mathrm{H}}\boldsymbol{G}_{A_k}^{\mathrm{H}}\boldsymbol{y}_k\|^2\big] \overset{(a)}{=} \|\boldsymbol{R}_{\boldsymbol{y}_k}^{\frac{1}{2}}(\widehat{\boldsymbol{G}}_k - \boldsymbol{G}_{A_k}\boldsymbol{G}_{D_k})\|_F^2,$$

$$\text{s.t.} \quad |[\boldsymbol{G}_{A_k}]_{n,m}| = 1, \tag{43}$$

Algorithm 2 Low-complexity BD-ZF/SLNR-Max hybrid transceiver designs

Input: BD-ZF/SLNR-Max fully-digital precoder $\boldsymbol{F}_k^{\mathrm{ZF}}/\boldsymbol{F}_k^{\mathrm{SL}}$, $\forall k$, derived from (31)/(50); initial analog precoder $\boldsymbol{F}_{A_k}^{(0)}$ and combiner $\boldsymbol{G}_{A_k}^{(0)}$, $\forall k$, derived from (30); outer iteration indexed $I_t = 0$ and $I_r = 0$.
1: Fix $\boldsymbol{F}_{A_k}^{(I_t)}$, $\forall k$, calculate $\boldsymbol{F}_{D_k}^{(I_t+1)}$, $\forall k$, according to (41).
2: Fix $\boldsymbol{F}_{D_k}^{(I_t+1)}$, $\forall k$, calculate $\boldsymbol{F}_{A_k}^{(I_t)}$, $\forall k$, via iteratively solving problem (36).
3: Set $I_t = I_t + 1$.
4: Fix $\boldsymbol{F}_{A_k}^{(I_t)}$ and $\boldsymbol{F}_{D_k}^{(I_t)}$, $\forall k$, calculate fully-digital combiner $\widehat{\boldsymbol{G}}_k$, $\forall k$.
5: **repeat**
6:     Fix $\boldsymbol{G}_{A_k}^{(I_r)}$, $\forall k$, calculate $\boldsymbol{G}_{D_k}^{(I_r+1)}$, $\forall k$, according to (44).
7:     Fix $\boldsymbol{G}_{D_k}^{(I_r+1)}$, $\forall k$, calculate $\boldsymbol{G}_{A_k}^{(I_r)}$, $\forall k$, using MM method as in Algorithm 1.
8:     Set $I_r = I_r + 1$.
9: **until** Objective function value of problem (43) converges.
Output: $\big\{\boldsymbol{G}_{D_k}^{(I_r)}, \boldsymbol{G}_{A_k}^{(I_r)}, \widehat{\boldsymbol{F}}_{D_k}^{(I_t)}, \boldsymbol{F}_{A_k}^{(I_t)}, \forall k\big\}$.

---

where $\boldsymbol{R}_{\boldsymbol{y}_k} = \mathbb{E}[\boldsymbol{y}_k\boldsymbol{y}_k^{\mathrm{H}}] = \sum_{i=1}^{K}\boldsymbol{H}_{k,i}\boldsymbol{F}_{A_i}\boldsymbol{F}_{D_i}\boldsymbol{F}_{D_i}^{\mathrm{H}}\boldsymbol{F}_{A_i}^{\mathrm{H}}\boldsymbol{H}_{k,i}^{\mathrm{H}} + \sigma_{n_k}^2\boldsymbol{I}_{N_{r_k}}$, and $\widehat{\boldsymbol{G}}_k = \boldsymbol{R}_{\boldsymbol{y}_k}^{-1}\boldsymbol{H}_{k,k}\boldsymbol{F}_{A_k}\boldsymbol{F}_{D_k}$. The equality (a) holds by following the similar derivations in [7]. Similarly, the proposed MM-based alternating optimization is still applicable to problem (43), which is more complicated than problem (33) due to the extra $\boldsymbol{R}_{\boldsymbol{y}_k}$. Firstly, when the analog combiner $\boldsymbol{G}_{A_k}$ is fixed, the optimal digital combiner $\boldsymbol{G}_{D_k}$ has the following closed-form expression

$$\boldsymbol{G}_{D_k} = \big(\boldsymbol{G}_{A_k}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{y}_k}\boldsymbol{G}_{A_k}\big)^{-1}\boldsymbol{G}_{A_k}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{y}_k}\widehat{\boldsymbol{G}}_k. \tag{44}$$

Based on the obtained $\boldsymbol{G}_{D_k}$ in (44), the MM method is then applied to tackle the nonconvex problem (43) w.r.t $\boldsymbol{G}_{A_k}$ by finding an appropriate majorized problem, which is

$$\min_{\boldsymbol{g}_{A_k}} \ \Re\{\boldsymbol{g}_{A_k}^{\mathrm{H}}\widetilde{\boldsymbol{r}}_k\}, \text{s.t.} \ |[\boldsymbol{g}_{A_k}]_n| = 1, \forall n=1,\cdots, N_{t_k}N_{t_k}^{RF}, \tag{45}$$

where $\widetilde{\boldsymbol{r}}_k = \big(\widetilde{\boldsymbol{R}}_k - \lambda_{\max}(\widetilde{\boldsymbol{R}}_k)\boldsymbol{I}_{N_{r_k}N_{r_k}^{RF}}\big)\boldsymbol{g}_{A_k}^{(l)} - \boldsymbol{r}_k$, $\boldsymbol{r}_k^{\mathrm{H}} = \mathrm{vec}\big((\boldsymbol{G}_{D_k}\widehat{\boldsymbol{G}}_k^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{y}_k})^{\mathrm{T}}\big)^{\mathrm{T}}$ and $\widetilde{\boldsymbol{R}}_k = \big(\boldsymbol{G}_{D_k}\boldsymbol{G}_{D_k}^{\mathrm{H}}\big)^{\mathrm{T}} \otimes \boldsymbol{R}_{\boldsymbol{y}_k}$. The semi closed-form solution to problem (45) is given by

$$\boldsymbol{g}_{A_k} = \mathrm{vec}(\boldsymbol{G}_{A_k}) = -e^{\mathrm{j}\arg(\widetilde{\boldsymbol{r}}_k)}. \tag{46}$$

Due to the iterative nature between (44) and (46), the obtained hybrid combiner better matches with the iterative-PP hybrid precoder in (37) and (41). Further, by integrating the sparse recovery problems (33) and (43), the proposed BD-ZF hybrid transceiver design is summarized in Algorithm 2.

B. SLNR-Max Hybrid Transceiver Design

It is known that the drawbacks of the BD-ZF technique are the restriction on the number of antennas and the noise enhancement. We consider an alternative design based on SLNR maximization. The SLNR of the kth transmitter is defined as the ratio of the received signal power at the desired kth receiver to the interference (leakage) at other receivers plus noise power [31]

$$\mathrm{SLNR}_k = \frac{\mathrm{Tr}\big((\boldsymbol{F}_k^{\mathrm{SL}})^{\mathrm{H}}\boldsymbol{H}_{k,k}^{\mathrm{H}}\boldsymbol{H}_{k,k}\boldsymbol{F}_k^{\mathrm{SL}}\big)}{\mathrm{Tr}\big(\sigma_{n_k}^2\boldsymbol{I}_{N_t^{RF}} + \sum_{i\neq k}(\boldsymbol{F}_k^{\mathrm{SL}})^{\mathrm{H}}\boldsymbol{H}_{k,i}^{\mathrm{H}}\boldsymbol{H}_{k,i}\boldsymbol{F}_k^{\mathrm{SL}}\big)}. \tag{47}$$

Then the SLNR-Max fully-digital precoder for each transmit-receiver pair is designed such that

$$\boldsymbol{F}_k^{\text{SL}} = \arg\max \text{SLNR}_k, \text{ s.t. } \text{Tr}\big((\boldsymbol{F}_k^{\text{SL}})^{\text{H}}\boldsymbol{F}_k^{\text{SL}}\big) \leq P_k, \forall k. \quad (48)$$

By defining the generalized EVD for the matrix pencil $\big(\boldsymbol{H}_{k,k}^{\text{H}}\boldsymbol{H}_{k,k}, \frac{N_{r_k}\sigma_{n_k}^2}{P_k}\boldsymbol{I}_{N_{t_k}^{RF}} + \sum_{i\neq k}\boldsymbol{H}_{k,i}^{\text{H}}\boldsymbol{H}_{k,i}\big)$ as

$$\begin{cases} \boldsymbol{T}_k^{\text{H}}\boldsymbol{H}_{k,k}^{\text{H}}\boldsymbol{H}_{k,k}\boldsymbol{T}_k = \boldsymbol{\Sigma}_k = \text{diag}[\sigma_{k,1}, \cdots, \sigma_{k,N_{t_k}^{RF}}], \\ \boldsymbol{T}_k^{\text{H}}\big(\frac{N_{r_k}\sigma_{n_k}^2}{P_k}\boldsymbol{I}_{N_{t_k}^{RF}} + \sum_{i\neq k}\boldsymbol{H}_{k,i}^{\text{H}}\boldsymbol{H}_{k,i}\big)\boldsymbol{T}_k = \boldsymbol{I}_{N_{t_k}^{RF}}, \forall k, \end{cases} \quad (49)$$

where the columns of $\boldsymbol{T}_k \in \mathbb{C}^{N_{t_k}^{RF}\times N_{t_k}^{RF}}$ and the diagonal elements of $\boldsymbol{\Sigma}_k$ are the generalized eigenvectors and eigenvalues, respectively, the optimal SLNR-Max fully-digital precoder is given by

$$\boldsymbol{F}_k^{\text{SL}} = \sqrt{\frac{P_k}{\text{Tr}\big(\boldsymbol{T}_k^{\text{H}}(:, 1:N_{s_k})\boldsymbol{T}_k(:, 1:N_{s_k})\big)}}\boldsymbol{T}_k(:, 1:N_{s_k}). \quad (50)$$

Similarly to the BD-ZF hybrid design, we formulate the SLNR-Max hybrid design by minimizing the Euclidean distance between $\boldsymbol{F}_k^{\text{SL}}$ in (50) and the hybrid precoder as

$$\begin{aligned} \min_{\boldsymbol{F}_{A_k}, \widetilde{\boldsymbol{F}}_{D_k}} \quad & \|\boldsymbol{F}_k^{\text{SL}} - \boldsymbol{F}_{A_k}\boldsymbol{F}_{D_k}\|_F^2, \\ \text{s.t.} \quad & |[\boldsymbol{F}_{A_k}]_{n,m}| = 1, \|\boldsymbol{F}_{A_k}\boldsymbol{F}_{D_k}\|_F^2 = P_k, \forall n, m, k. \end{aligned} \quad (51)$$

Notice that problem (51) can be effectively addressed following the same approach of solving problem (33) by replacing $\boldsymbol{F}_k^{\text{ZF}}$ with $\boldsymbol{F}_k^{\text{SL}}$, implying that the MM-based alternating optimization is still applicable. Additionally, once the SLNR-Max hybrid precoder for each transmit-receive pair is obtained, the corresponding MMSE hybrid combiner design can be independently carried out as (43). This SLNR-Max hybrid design is also summarized in Algorithm 2.

C. Partially-Connected Hybrid Transceiver Structure

In the partially-connected structure, each RF chain at both ends is only connected with a part of the antenna array. Specifically, at the kth transmitter (receiver), each RF chain is only connected with $N_{t_k}/N_{t_k}^{RF}$ ($N_{r_k}/N_{r_k}^{RF}$) antennas, and thus the analog precoder $\boldsymbol{F}_{A_k}$ and combiner $\boldsymbol{G}_{A_k}$, $\forall k$, can be expressed by the following block matrices

$$\begin{aligned} \boldsymbol{F}_{A_k} &= \text{BLkdiag}[\boldsymbol{p}_{k_1}, \boldsymbol{p}_{k_2} \cdots \boldsymbol{p}_{k_{N_{t_k}^{RF}}}], \\ \boldsymbol{G}_{A_k} &= \text{BLkdiag}[\boldsymbol{q}_{k_1}, \boldsymbol{q}_{k_2} \cdots \boldsymbol{q}_{k_{N_{r_k}^{RF}}}], \end{aligned} \quad (52)$$

where the unit-modulus entries $|[\boldsymbol{p}_{i_k}]_{m_k}| = 1$, $\forall i_k = 1, \cdots, N_{t_k}^{RF}$, $\forall m_k = 1, \cdots, N_{t_k}/N_{t_k}^{RF}$, and $|[\boldsymbol{q}_{j_k}]_{n_k}| = 1$, $\forall j_k = 1, \cdots, N_{r_k}^{RF}$, $\forall n_k = 1, \cdots, N_{r_k}/N_{r_k}^{RF}$, are imposed. Benefited from the block diagonal structures of the analog precoder and combiner, the MM-based alternating optimization can be directly applied to the WMMSE problem (6) to obtain the locally optimal solution without requiring the approximation on analog precoder as in Section III. More importantly, due to the sparsity of the partially-connected structure, the MM-based analog precoder and combiner designs exhibit much lower complexity than that of Section III.

1) *Semi Closed-Form Digital Precoder* $\boldsymbol{F}_{D_k}^{\text{Par}}$: Based on the partially-connected structure (52), we can re-express $\widetilde{\boldsymbol{F}}_{A_k}$ and $\widetilde{\boldsymbol{F}}_{D_k}$ as $\widetilde{\boldsymbol{F}}_{A_k} = \sqrt{N_{t_k}^{RF}/N_{t_k}}\boldsymbol{F}_{A_k}$ and $\widetilde{\boldsymbol{F}}_{D_k} = \sqrt{N_{t_k}/N_{t_k}^{RF}}\boldsymbol{F}_{D_k}$, respectively, which are then substituted into (16) to obtain the semi closed-form digital precoder $\boldsymbol{F}_{D_k}^{\text{Par}}$

$$\boldsymbol{F}_{D_k}^{\text{Par}} = \frac{1}{\sqrt{N_{t_k}}}\big(\sum_{i=1}^K \boldsymbol{L}_{i,k}^{\text{H}}\boldsymbol{W}_i\boldsymbol{L}_{i,k} + \beta_k'\boldsymbol{I}_{N_{t_k}^{RF}}\big)^{-1}\boldsymbol{L}_{k,k}^{\text{H}}\boldsymbol{W}_k. \quad (53)$$

where the determination of scalar β_k' is similar to β_k.

2) *Semi Closed-Form Analog Precoder* $\boldsymbol{F}_{A_k}^{\text{Par}}$: Given $\{\boldsymbol{F}_{D_k}^{\text{Par}}, \forall k\}$, we firstly define the following auxiliary parameters for optimizing the partially-connected analog precoder $\boldsymbol{F}_{A_k}^{\text{Par}}$, which are

$$\begin{cases} \widehat{\boldsymbol{A}}_k = \frac{N_{t_k}^{RF}}{N_{t_k}}\begin{bmatrix} \widehat{\boldsymbol{A}}_k^{1,1} & \cdots & \widehat{\boldsymbol{A}}_k^{1,N_{t_k}^{RF}} \\ \vdots & \ddots & \vdots \\ \widehat{\boldsymbol{A}}_k^{N_{t_k}^{RF},1} & \cdots & \widehat{\boldsymbol{A}}_k^{N_{t_k}^{RF},N_{t_k}^{RF}} \end{bmatrix} \in \mathbb{C}^{N_{t_k}\times N_{t_k}}, \\ \widehat{\boldsymbol{A}}_k^{l,q} = \widetilde{\boldsymbol{A}}_k\big(\widetilde{l}:\widetilde{l}+\frac{N_{t_k}}{N_{t_k}^{RF}}-1, \widetilde{q}:\widetilde{q}+\frac{N_{t_k}}{N_{t_k}^{RF}}-1\big), \\ \bar{\boldsymbol{f}}_{A_k} = [\boldsymbol{p}_1^{\text{T}}\cdots\boldsymbol{p}_{N_{t_k}^{RF}}^{\text{T}}]^{\text{T}} \in \mathbb{C}^{N_{t_k}}, \widehat{\boldsymbol{a}}_k = [\widehat{\boldsymbol{a}}_k^1\cdots\widehat{\boldsymbol{a}}_k^{N_{t_k}^{RF}}] \in \mathbb{C}^{N_{t_k}}, \\ \widehat{\boldsymbol{a}}_k^l = \sqrt{\frac{N_{t_k}^{RF}}{N_{t_k}}}\boldsymbol{a}_k\big(\widetilde{l}:\widetilde{l}+\frac{N_{t_k}}{N_{t_k}^{RF}}-1\big), \widetilde{l} = (l-1)\big(\frac{N_{t_k}}{N_{t_k}^{RF}}+N_{t_k}\big)+1, \\ \widetilde{q} = (q-1)\big(\frac{N_{t_k}}{N_{t_k}^{RF}}+N_{t_k}\big)+1, \forall l, q = 1, \cdots, N_{t_k}^{RF}. \end{cases} \quad (54)$$

Following the similar derivations of (19), the partially-connected analog precoder $\boldsymbol{F}_{A_k}^{\text{Par}}$ for each transmit-receive pair is independently designed as

$$\begin{aligned} \min_{\bar{\boldsymbol{f}}_{A_k}} \quad & \bar{\boldsymbol{f}}_{A_k}^{\text{H}}\widehat{\boldsymbol{A}}_k\bar{\boldsymbol{f}}_{A_k} - 2\Re\{\widehat{\boldsymbol{a}}_k^{\text{H}}\bar{\boldsymbol{f}}_{A_k}\}, \\ \text{s.t.} \quad & |[\bar{\boldsymbol{f}}_{A_k}]_n| = 1, n = 1, \cdots, N_{t_k}. \end{aligned} \quad (55)$$

Recalling Lemma 1, the majorized counterpart of the problem (55) at $\bar{\boldsymbol{f}}_{A_k}^{(l)}$ is formulated as

$$\min_{\bar{\boldsymbol{f}}_{A_k}} \Re\{\vec{\boldsymbol{a}}_k^{\text{H}}\bar{\boldsymbol{f}}_{A_k}\}, \text{ s.t } |[\bar{\boldsymbol{f}}_{A_k}]_n| = 1, \forall n = 1, \cdots, N_{t_k}, \quad (56)$$

where $\vec{\boldsymbol{a}}_k = \big(\widehat{\boldsymbol{A}}_k - \lambda_{\max}(\widehat{\boldsymbol{A}}_k)\boldsymbol{I}_{N_{t_k}}\big)\widetilde{\boldsymbol{f}}_{A_k}^{(l)} - \widehat{\boldsymbol{a}}_k$, and the semi closed-form solution is obtained as

$$\bar{\boldsymbol{f}}_{A_k} = \text{vec}(\boldsymbol{F}_{A_k}^{\text{Par}}) = -e^{\text{j}\arg(\vec{\boldsymbol{a}}_k)}. \quad (57)$$

3) *Semi Closed-Form Analog Combiner* $\boldsymbol{G}_{A_k}^{\text{Par}}$: Similarly to solving (23), by defining

$$\begin{cases} \widehat{\boldsymbol{N}}_k = \begin{bmatrix} \widehat{\boldsymbol{N}}_k^{1,1} & \cdots & \widehat{\boldsymbol{N}}_k^{1,N_{r_k}^{RF}} \\ \vdots & \ddots & \vdots \\ \widehat{\boldsymbol{N}}_k^{N_{r_k}^{RF},1} & \cdots & \widehat{\boldsymbol{N}}_k^{N_{r_k}^{RF},N_{r_k}^{RF}} \end{bmatrix} \in \mathbb{C}^{N_{r_k}\times N_{r_k}}, \\ \widehat{\boldsymbol{N}}_k^{l,q} = \widetilde{\boldsymbol{N}}_k\big(\widetilde{l}:\widetilde{l}+\frac{N_{r_k}}{N_{r_k}^{RF}}-1, \widetilde{q}:\widetilde{q}+\frac{N_{r_k}}{N_{r_k}^{RF}}-1\big), \\ \bar{\boldsymbol{g}}_{A_k} = [\boldsymbol{q}_1^{\text{T}}\cdots\boldsymbol{q}_{N_{r_k}^{RF}}^{\text{T}}]^{\text{T}} \in \mathbb{C}^{N_{r_k}}, \widehat{\boldsymbol{d}}_k = [\widehat{\boldsymbol{d}}_k^1\cdots\widehat{\boldsymbol{d}}_k^{N_{r_k}^{RF}}] \in \mathbb{C}^{N_{r_k}}, \\ \widehat{\boldsymbol{d}}_k^l = \boldsymbol{d}_k\big(\widetilde{l}:\widetilde{l}+\frac{N_{r_k}}{N_{r_k}^{RF}}-1\big), \widetilde{l} = (l-1)\big(\frac{N_{r_k}}{N_{r_k}^{RF}}+N_{r_k}\big)+1, \\ \widetilde{q} = (q-1)\big(\frac{N_{r_k}}{N_{r_k}^{RF}}+N_{r_k}\big)+1, \forall l, q = 1, \cdots, N_{r_k}^{RF}, \end{cases} \quad (58)$$

the independent design of partially-connected analog combiner $\boldsymbol{G}_{A_k}^{\mathrm{Par}}$ for each transmit-receive pair can be formulated as

$$\min_{\bar{\boldsymbol{g}}_{A_k}} \ \bar{\boldsymbol{g}}_{A_k}^{\mathrm{H}} \widehat{\boldsymbol{N}}_k \bar{\boldsymbol{g}}_{A_k} - 2\Re\big\{\widehat{\boldsymbol{d}}_k^{\mathrm{H}} \bar{\boldsymbol{g}}_{A_k}\big\},$$
$$\text{s.t. } |[\bar{\boldsymbol{g}}_{A_k}]_n| = 1, \forall n = 1, \cdots, N_{r_k}. \quad (59)$$

Also, the majorized counterpart of the problem (59) at $\bar{\boldsymbol{g}}_{A_k}^{(l)}$ can be expressed as

$$\min_{\bar{\boldsymbol{g}}_{A_k}} \ \Re\big\{\bar{\boldsymbol{d}}_k^{\mathrm{H}} \bar{\boldsymbol{g}}_{A_k}\big\}, \text{ s.t } |[\bar{\boldsymbol{g}}_{A_k}]_n| = 1, \forall n = 1, \cdots, N_{r_k}. \quad (60)$$

where $\bar{\boldsymbol{d}}_k = \big(\widehat{\boldsymbol{N}}_k - \lambda_{\max}(\widehat{\boldsymbol{N}}_k)\boldsymbol{I}_{N_{r_k}}\big)\bar{\boldsymbol{g}}_{A_k}^{(l)} - \widehat{\boldsymbol{d}}_k$, and the semi closed-form solution is derived as

$$\bar{\boldsymbol{g}}_{A_k} = \mathrm{vec}\big(\boldsymbol{G}_{A_k}^{\mathrm{Par}}\big) = -e^{\mathrm{j}\arg\big(\bar{\boldsymbol{d}}_k\big)}, \ \forall k. \quad (61)$$

4) Semi Closed-Form Digital Combiner $\boldsymbol{G}_{D_k}^{\mathrm{Par}}$ and weighting matrix $\boldsymbol{W}_k^{\mathrm{Par}}$: The optimal digital combiner $\boldsymbol{G}_{D_k}$ for the WMMSE problem (6) under this partially-connected structure is also Wiener filter, which has the same form as (10). Moreover, the optimal weighing matrix $\boldsymbol{W}_k^{\mathrm{Par}}$ can be similarly derived as (9).

Observing from (56) and (60) that this partially-connected structure simplifies the analog precoder and combiner design due to the reduced number of optimization variables, and also makes the proposed MM-based alternating optimization directly applicable without the assumption in large-scale MIMO regime. In a nutshell, the proposed MM-based hybrid design is well suited for this partially-connected structure.

V. Convergence and Complexity Analysis of Proposed Algorithms

In this section, we firstly study the convergence of the proposed MM-based alternating optimization for solving problem (8), as elaborated below.

Proposition 4. *Every limit point $(\boldsymbol{W}_k^{\infty}, \boldsymbol{G}_{A_k}^{\infty}, \boldsymbol{G}_{D_k}^{\infty}, \boldsymbol{F}_{A_k}^{\infty}, \widetilde{\boldsymbol{F}}_{D_k}^{\infty}, \forall k)$ of the iterates generated by the proposed MM-Alt-Opt is a (Boulingand) stationary point of the approximated WMMSE problem (8) and also a high-quality suboptimal solution of the original WMMSE problem (6). In particular, when considering the mmWave K-user MIMO interference system, the obtained $(\boldsymbol{W}_k^{\infty}, \boldsymbol{G}_{D_k}^{\infty}, \boldsymbol{G}_{A_k}^{\infty}, \boldsymbol{F}_{A_k}^{\infty}, \widetilde{\boldsymbol{F}}_{D_k}^{\infty}, \forall k)$ is an asymptotically stationary point to problem (6) as $N_t \to \infty$.*

Proof. See Appendix D. □

Next, we analyze the computational complexity of the proposed MM-Alt-Opt, PP-based two-stage hybrid design (Hybrid PP-Two-Stage), BD-ZF and SLNR-Max based hybrid designs (Hybrid BD-ZF/SLNR-Max), in comparison with the classical OMP scheme [7]. To simplify the analysis, we consider that $N_t = N_{t_k}$, $N_r = N_{r_k}$ and $N_{RF} = N_{t_k}^{RF} = N_{r_k}^{RF} = N_{s_k}, \forall k$. In the OMP scheme, the length of codebooks for analog precoder (combiner) design is set to L_c, and $N_t > N_r \gg L_c > N_{RF}$ is assumed. We focus on the complexity of major computational steps with the low-order terms omitted, and then the total complexity is derived.
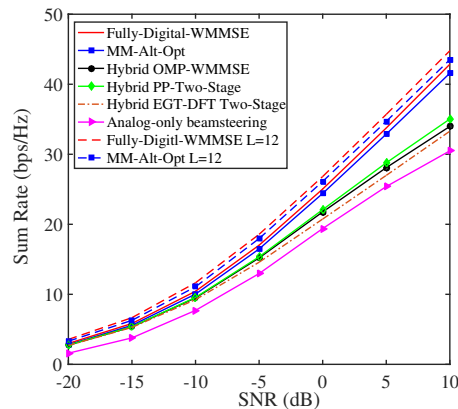

Fig. 2. Sum rate performance versus SNR in the mmWave channel achieved by the proposed MM-Alt-Opt and Hybrid PP-Two-Stage as well as the benchmarks Hybrid OMP-WMMSE, Hybrid EGT-DFT Two-Stage and Analog-only beamsteering, using the Fully-Digital-WMMSE as the upper bound. The sum rates of the MM-Alt-Opt and Fully-Digital-WMMSE for the $L=12$ scatters are also shown.

Let I_W and I_M be the numbers of outer and inner iterations, respectively, for the MM-based methods, including the MM-Alt-Opt and Hybrid BD-ZF/SLNR-Max. Observe from Algorithm 1 that in one outer iteration of the MM-Alt-Opt, the computational cost is mainly from the MM-based analog precoder design with the complexity on the order of $\mathsf{O}(I_M N_t^2 N_{RF}^2)$ per transmit-receive pair. The total complexity of the MM-Alt-Opt is obviously linear w.r.t. the number of outer iterations I_W and the number of communication pairs K. The similar analysis is applicable to the partially-connected hybrid transceiver case (Hybrid ParTxRx). While for the Hybrid PP-Two-Stage, the complexity primarily comes from the selection of analog precoder and combiner based on the SVD of $N_r \times N_t$ channel matrix for each transmit-receive pair. The complexity of designing $\boldsymbol{F}_{D_k}$ and $\boldsymbol{G}_{D_k}$, which involves an iterative loop with I_O iterations, is much smaller by comparison. This yields the total complexity of $\mathsf{O}(K N_t^2 N_r)$. For the Hybrid BD-ZF/SLNR-Max, by defining I_P as the number of iterations for the iterative-PP method, the hybrid precoder design has the complexity $\mathsf{O}(K I_P N_t^2 N_{RF})$, while the MM-based analog combiner design has the complexity $\mathsf{O}(K I_W I_M N_r^2 N_{RF}^2)$. Hence the total complexity of this scheme is $\mathsf{O}(K I_P N_t^2 N_{RF})$ for the large N_t. The OMP scheme involves an exhaustive search for both analog precoder and combiner from the predefined codebooks and large-scale matrix multiplication, yielding the total complexity $\mathsf{O}(K I_B N_t^3)$, where I_B is the number of iterations for finding the WMMSE digital precoder and combiner.

VI. Simulation Results

Unless otherwise stated, $K = 2$ transceiver pairs are used. Each transmitter deploys $N_t = 64$ antennas with $N_t^{RF} = 4$ RF chains to send $N_s = 4$ data streams to its receiver, which has $N_r = 16$ antennas and $N_t^{RF} = 4$ RF chains. The RF phase shifters with infinite resolution are assumed. Both the Rayleigh and mmWave channels are considered. For the normalized Rayleigh channel, the elements of all channel matrices are distributed according to $\mathcal{CN}(0, 1)$. For the normalized mmWave channel, the propagation environment with $L_k = L = 10$ scatters, $\forall k$, is considered, in which the
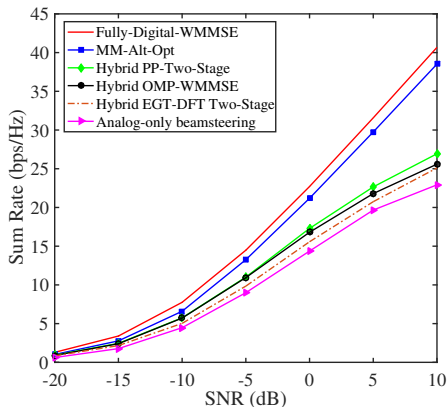
Fig. 3. Sum rate performance versus SNR in the Rayleigh channel achieved by the proposed MM-Alt-Opt and Hybrid PP-Two-Stage as well as the benchmarks Hybrid OMP-WMMSE, Hybrid EGT-DFT Two-Stage and Analog-only beamsteering, using the Fully-Digital-WMMSE as the upper bound.
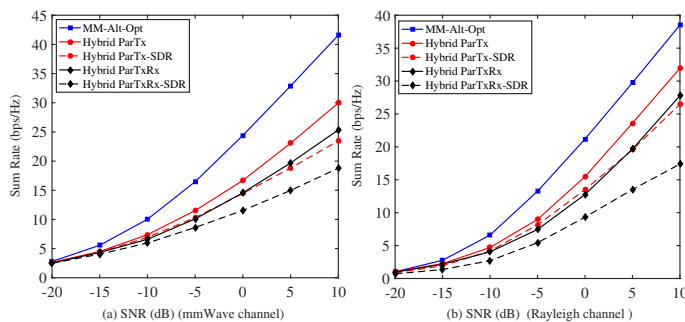

Fig. 4. Sum rate performance versus SNR in (a) the mmWave channel and (b) the Rayleigh channel, achieved by the proposed MM-Alt-Opt, Hybrid-ParTx and Hybrid-ParTxRx as well as the benchmarks Hybrid ParTx-SDR and Hybrid ParTxRx-SDR.

AOA and AOD of each path are uniformly distributed in $[0, 2\pi]$, while the pathloss factors $\alpha_k = \alpha$, $\forall k$, with α obeying $\mathcal{CN}(0, 1)$. By assuming the same transmit power $P_k = P$ and the same noise power $\sigma_{n_k}^2 = \sigma_n^2$ at all transmitters and receivers, respectively, the received SNR becomes $\mathrm{SNR} = \frac{P}{\sigma_n^2}$. All the results are obtained by averaging over 100 channel realizations.

In this work, we propose various hybrid transceiver designs, including the **MM-Alt-Opt**, the **Hybrid PP-Two-Stage**, the **Hybrid BD-ZF/SLNR-Max**, and the partially-connected hybrid structure of **Hybrid-ParTxRx**. In fact, there is another scheme which only considers the partially-connected hybrid structure at transmitter, and we call this scheme **Hybrid-ParTx**. The sum rate performance of these proposed designs are compared with that of the following baselines:

Hybrid OMP [7]: The sparse reconstruction of the hybrid precoder and combiner of each transmit-receive pair is realized from the fully-digital precoder and MMSE combiner as well as predetermined codebook . The analog beamforming codebook used consists of the array steering vectors (the left/right singular vectors with phase mapping) of the desired mmWave (Rayleigh) channel. In particular, three baselines, called **Hybrid OMP-WMMSE**, **Hybrid OMP-ZF** and **Hybrid OMP-SLNR**, are adopted according to three different fully-digital precoders based on the WMMSE, BD-ZF and SLNR-Max criteria, respectively. **Hybrid EGT-DFT Two-**
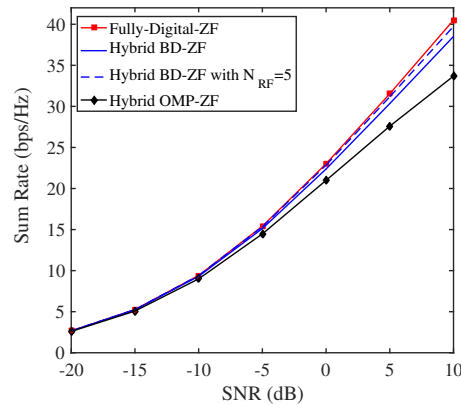

Fig. 5. Sum rate performance versus SNR in the mmWave channel achieved by the proposed Hybrid BD-ZF and the benchmark Hybrid OMP-ZF, using the Fully-Digital-ZF as the upper bound. The sum rate of the Hybrid BD-ZF with $N_t^{RF} = N_r^{RF} = N_{RF} = 5$ is also shown.

Stage [16]: The EGT based analog precoder and DFT based analog combiner harvest the large array gain in the first analog stage, and the inter-user interference elimination is left to the second digital stage. **Hybrid ParTx-SDR/Hybrid ParTxRx-SDR** [19]: First the Euclidean distance between the partially-connected hybrid precoder and the fully-digital WMMSE precoder is minimized in which the iterative procedure between the semidefinite relaxation (SDR) based digital precoder and the PP-based analog precoder is performed. Then the MM-based hybrid combiner designs under the fully connected and partially-connected receiver structures are performed, corresponding to **Hybrid ParTx-SDR** and **Hybrid ParTxRx-SDR**, respectively. **Analog-only beamsteering** [27]: Only analog beamforming strategies at both ends are considered to align transmit and receive beams of each transceiver pair for maximizing array gain. The inter-user interference elimination is not involved.

Furthermore, the near-optimal fully-digital schemes based on the criteria of WMMSE, BD-ZF and SLNR-Max (**Fully-Digital-WMMSE**, **Fully-Digital-ZF** and **Fully-Digital-SLNR**) are adopted as the corresponding upper-bound benchmarks.

Fig. 2 compares the sum rate performance versus SNR in the mmWave channel achieved by the MM-Alt-Opt and Hybrid PP-Two-Stage with those of the three benchmarks, using the Fully-Digital-WMMSE as the upper bound. It can be seen from Fig. 2 that the sum rate of our MM-Alt-Opt is very close to the optimal Fully-Digital-WMMSE, confirming that it is near-optimal. Benefited from its iterative nature, the MM-Alt-Opt clearly outperforms the Hybrid PP-Two-Stage with one-shot approximation for analog precoder and combiner design. Also the Hybrid PP-Two-Stage achieves a similar performance to the Hybrid OMP-WMMSE at low SNR region, but slightly better performance at high SNR region. More importantly, the Hybrid PP-Two-Stage does not require the WMMSE fully-digital solution and has much lower-complexity than the Hybrid OMP-WMMSE. Since the inter-user interference elimination is not considered in the Analog-only beamsteering, its performance is the worst. In addition, when a larger number of scatters is considered, i.e., $L = 12$, the MM-Alt-Opt still
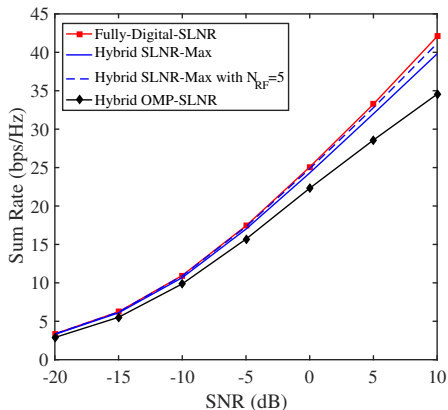
Fig. 6. Sum rate performance versus SNR in the mmWave channel achieved by the proposed Hybrid SLNR-Max and the benchmark Hybrid OMP-SLNR, using the Fully-Digital-SLNR as the upper bound. The sum rate of the Hybrid SLNR-Max with $N_t^{RF} = N_r^{RF} = N_{RF} = 5$ is also shown.
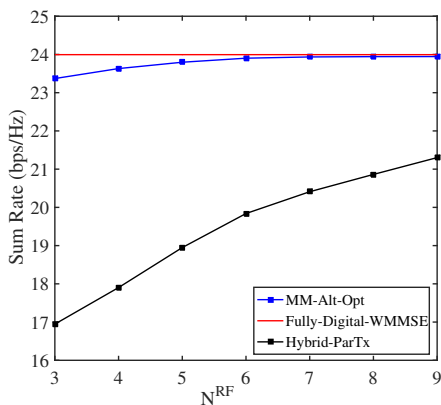

Fig. 7. Sum rate performance versus number of transmit/receive RF chains $N_t^{RF} = N_r^{RF} = N^{RF}$ in the mmWave channel achieved by the proposed MM-Alt-Opt and Hybrid-ParTx, in comparison with the Fully-Digital-WMMSE, given SNR $= 0$ dB.

performs almost as good as the Fully-Digital-WMMSE, both having slightly higher sum rate compared to the case of $L = 10$.

Next, we carry out the same comparison in the Rayleigh scenario and the results are shown in Fig. 3, where the performance gap between the optimal Fully-Digital-WMMSE and the MM-Alt-Opt is larger than that in the mmWave channel. The reason is that the approximation $\boldsymbol{F}_{A_k}^{\mathrm{H}} \boldsymbol{F}_{A_k} \approx N_t \boldsymbol{I}_{N_t^{RF}}$ adopted in the MM-Alt-Opt is less accurate for the Rayleigh fading channel.

From Fig. 4, it can be seen that the MM-Alt-Opt considerably outperforms the Hybrid-ParTx in both the mmWave and Rayleigh fading cases, since the inter-user interference cannot be effectively suppressed by the Hybrid-ParTx with much reduced design freedom in analog precoder. Similarly, the Hybrid-ParTx has better sum rate performance than the Hybrid-ParTxRx, since the latter has the further much reduced design freedom in analog combiner. Also, observe from Fig. 4 that the proposed Hybrid-ParTx outperforms its corresponding benchmark Hybrid ParTx-SDR, while the Hybrid-ParTxRx outperforms its related baseline Hybrid ParTxRx-SDR.

Fig. 5 compares the sum rate performance of the proposed Hybrid BD-ZF and the baseline Hybrid OMP-ZF, using the Fully-Digital-ZF solution as the upper bound. Observe that the sum rate of the Hybrid BD-ZF is close to that of the full-digital BD-ZF solution, especially when one extra RF, i.e.,
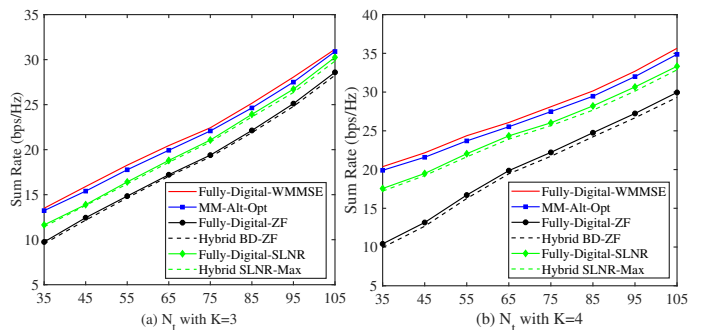

Fig. 8. Sum rate performance versus the number of transmit antennas N_t in the mmWave channel achieved by the proposed MM-Alt-Opt, Hybrid BD-ZF and Hybrid SLNR-Max, in comparison with the corresponding optimal Fully-Digital-WMMSE, Fully-Digital-ZF and Fully-Digital-SLNR: (a) $K = 3$ users, and (b) $K = 4$ users.

$N_{RF} = 5$, is considered. Moreover, the proposed hybrid BD-ZF clearly achieves higher sum rate than the hybrid OMP-ZF baseline, because its iterative nature enables the hybrid precoder to better approximate the fully-digital solution at the expense of higher computational complexity. Furthermore, Fig. 6 shows the sum rates achieved by the proposed Hybrid SLNR-Max and the baseline Hybrid-OMP-SLNR versus SNR in the mmWave channel, using the Fully-Digital-SLNR as the upper bound. Clearly, Fig. 6 presents similar comparison results among the three schemes to Fig. 5.

Fig. 7 compares the sum rate performance versus the number of RF chains N_{RF} in the mmWave channel achieved by the proposed MM-Alt-Opt and hybrid-ParTx as well as the optimal Fully-Digital-WMMSE. For the hybrid-ParTx, each of the first $N_t^{RF} - 1$ RF chains is connected with $\left\lfloor \frac{N_t}{N_t^{RF}} \right\rfloor$ transmit antennas, while the last RF chain is connected with $N_t - (N_t^{RF} - 1) \left\lfloor \frac{N_t}{N_t^{RF}} \right\rfloor$ antennas. It has been shown in [12] that when $N_{RF} \geq 2N_s$, there exists a globally optimal hybrid precoder and combiner design, which perfectly reconstructs the fully-digital precoder and combiner, yielding the same sum rate performance. Observe from Fig. 7 that almost identical performance are attained by both the Fully-Digital-WMMSE and the MM-Alt-Opt when $N_{RF} \geq 2N_s = 8$. Obviously, the Hybrid-ParTx cannot perfectly reconstruct the fully-digital design due to the reduced design freedom of analog precoder, and the achievable sum rate of the Hybrid-ParTx increases with N_{RF} mainly owing to the increased design freedom of digital precoder.

Finally, Fig. 8 depicts the sum rates as functions of the number of transmit antennas N_t in the mmWave channel achieved by the proposed MM-Alt-Opt, Hybrid BD-ZF and Hybrid SLNR-Max, in comparison to their corresponding optimal Fully-Digital-WMMSE, Fully-Digital-ZF and Fully-Digital-SLNR counterparts, respectively. Specifically, the numbers of transmit-receive pairs $K = 3$ and $K = 4$ are considered in Fig. 8 (a) and Fig. 8 (b), respectively. Both the results of Fig. 8 (a) and Fig. 8 (b) confirm that the proposed MM-Alt-Opt, Hybrid BD-ZF and Hybrid SLNR-Max perform close to their respective fully-digital counterparts. Naturally, the achievable sum rates of all studied schemes increase with N_t owing to the increased spatial DoFs. Moreover, the MM-Alt-Opt performs best in terms of the achievable sum rate, while the Hybrid BD-ZF scheme performs worst. When considering $K = 4$ in Fig. 8 (b), the performance advantage of the MM-

Alt-Opt over the other two proposed hybrid designs becomes more pronounced due to its higher design freedom.

VII. CONCLUSIONS

This paper has investigated various hybrid transceiver designs for sum rate maximization in both the mmWave and Rayleigh K-user MIMO interference channels. First, by-passing the near-optimal WMMSE fully-digital solution, we have jointly designed hybrid precoder and combiner in an alternating manner, in which the MM method is used to design the analog precoder and combiner. Moreover, a PP-based two-stage scheme has been proposed to decouple the design of analog and digital precoder (combiner), leading to lower complexity. Second, with the aid of the easy-to-implement fully-digital precoder, the low-complexity BD-ZF and SLNR-Max hybrid schemes have been studied, which focus on approximating the hybrid precoders to the fully-digital solutions derived according to the BD-ZF and SLNR criteria, respectively. Third, the partially-connected transceiver structure has been considered to reduce the system hardware cost and complexity, to which the MM-based alternating optimization is applicable. Numerical results have demonstrated the effectiveness of all our proposed hybrid transceiver designs, and they have shown that the sum rate performance of all our proposed hybrid designs are close or superior to those of the existing benchmarks. Our future research will study all the proposed hybrid designs implemented with finite resolution phase shifters and/or with limited channel feedback.

APPENDIX A

In fact, by regarding the analog precoders $\boldsymbol{F}_{A_k}$'s and combiners $\boldsymbol{G}_{A_k}$'s as a part of the effective channels between the ith transmitter and the kth receiver, i.e. $\boldsymbol{H}_{k,i}^{\mathrm{eff}} = \boldsymbol{G}_{A_k}\boldsymbol{H}_{k,i}\boldsymbol{F}_{A_i}, \forall i,k$, the equivalence between the sum-rate maximization problem (4) and the WMMSE problem (6) can still be established by substituting both the optimal weighting matrix and the optimal digital MMSE combiner of the WMMSE problem (6) into its objective function, where the constant N_S actually denotes an offset between the WMMSE w.r.t. $\boldsymbol{W}_k$, i.e. $\min \sum_{k=1}^{K} \mathrm{Tr}\big(\boldsymbol{W}_k\boldsymbol{E}_k(\mathcal{A}_k)\big) - \log\det(\boldsymbol{W}_k)$, and the achievable sum rate.

To be specific, for any given $\{\boldsymbol{G}_{D_k}, \boldsymbol{G}_{A_k}, \widetilde{\boldsymbol{F}}_{D_k}, \widetilde{\boldsymbol{F}}_{D_k}\}$, it is observed that the WMMSE problem (6) is convex w.r.t $\boldsymbol{W}_k$. Therefore, by setting the first-order derivative of its objective function w.r.t $\boldsymbol{W}_k$ to zero, we have the optimal $\boldsymbol{W}_k = \boldsymbol{E}_k(\mathcal{A}_k)^{-1}$. Substituting this optimal $\boldsymbol{W}_k$ into problem (6) yields the following equivalent counterpart

$$\max_{\mathcal{A}_k} \quad \sum_{k=1}^{K} \log\det\left(\big(\boldsymbol{E}_k(\mathcal{A}_k)\big)^{-1}\right),$$
$$\text{s.t. } \mathrm{Tr}\big(\widetilde{\boldsymbol{F}}_{D_k}^{\mathrm{H}}\widetilde{\boldsymbol{F}}_{D_k}\big) \le P_k, \ \widetilde{\boldsymbol{F}}_{A_k} = \boldsymbol{F}_{A_k}\big(\boldsymbol{F}_{A_k}^{\mathrm{H}}\boldsymbol{F}_{A_k}\big)^{-\frac{1}{2}},$$
$$\big|[\boldsymbol{F}_{A_k}]_{n,m}\big| = 1, \ \big|[\boldsymbol{G}_{A_k}]_{n,m}\big| = 1, \ \forall k,n,m, \quad (62)$$

Furthermore, to derive the unconstrained optimal $\boldsymbol{G}_{D_k}$ of problem (62), we recall the definition of the MSE matrix

$\boldsymbol{E}_k(\mathcal{A}_k)$ in (7) and define the following auxiliary parameters.

$$\boldsymbol{T}_k = \boldsymbol{G}_{A_k}^{\mathrm{H}}\boldsymbol{H}_{k,k}\widetilde{\boldsymbol{F}}_{A_k}\widetilde{\boldsymbol{F}}_{D_k}, \quad \boldsymbol{Q}_k = \Big(\sigma_{n_k}^2\boldsymbol{G}_{A_k}^{\mathrm{H}}\boldsymbol{G}_{A_k}+$$
$$\sum_{i=1}^{K}\big(\boldsymbol{G}_{A_k}^{\mathrm{H}}\boldsymbol{H}_{k,i}\widetilde{\boldsymbol{F}}_{A_i}\widetilde{\boldsymbol{F}}_{D_i}\big)\big(\boldsymbol{G}_{A_k}^{\mathrm{H}}\boldsymbol{H}_{k,i}\widetilde{\boldsymbol{F}}_{A_i}\widetilde{\boldsymbol{F}}_{D_i}\big)^{\mathrm{H}}\Big)^{-1}, \quad (63)$$

Based on (63), the MSE matrix $\boldsymbol{E}_k(\mathcal{A}_k)$ can be further expressed as

$$\boldsymbol{E}_k(\mathcal{A}_k) = \boldsymbol{G}_{D_k}^H\boldsymbol{Q}_k^{-1}\boldsymbol{G}_{D_k} - \boldsymbol{G}_{D_k}^H\boldsymbol{T}_k - \boldsymbol{T}_k^H\boldsymbol{G}_{D_k} + \boldsymbol{I}_{N_{s_k}}$$
$$= \big(\boldsymbol{G}_{D_k} - \boldsymbol{Q}_k\boldsymbol{T}_k\big)^H\boldsymbol{Q}_k^{-1}\big(\boldsymbol{G}_{D_k} - \boldsymbol{Q}_k\boldsymbol{T}_k\big) + \boldsymbol{I}_{N_{s_k}} - \boldsymbol{T}_k^H\boldsymbol{Q}_k\boldsymbol{T}_k$$
$$\succeq \boldsymbol{I}_{N_{s_k}} - \boldsymbol{T}_k^H\boldsymbol{Q}_k\boldsymbol{T}_k. \quad (64)$$

Using (64), we have $\log\det\big((\boldsymbol{E}_k(\mathcal{A}_k))^{-1}\big) \le \log\det\big((\boldsymbol{I}_{N_{s_k}} - \boldsymbol{T}_k^H\boldsymbol{Q}_k\boldsymbol{T}_k)^{-1}\big)$ and thus the optimal $\boldsymbol{G}_{D_k} = \boldsymbol{Q}_k\boldsymbol{T}_k$ of problem (62) can be obtained, based on which the objective function of the WMMSE problem (6) can be rewritten as

$$\sum_{k=1}^{K} \log\det\left(\big(\boldsymbol{E}_k(\mathcal{A}_k)\big)^{-1}\right)$$
$$= \sum_{k=1}^{K} \log\det\left(\big(\boldsymbol{I}_{N_{s_k}} - \boldsymbol{T}_k^H\boldsymbol{Q}_k\boldsymbol{T}_k\big)^{-1}\right), \quad (65a)$$

Furthermore, by defining $\widetilde{\boldsymbol{G}}_{D_k} = \boldsymbol{Q}_k^{-\frac{1}{2}}\boldsymbol{G}_{D_k}$, we also reexpress the achievable sum rate R_{sum} in problem (4) as

$$R_{\mathrm{sum}} = \sum_{k=1}^{K} \log\det(\boldsymbol{G}_{D_k}^H\boldsymbol{Q}_k^{-1}\boldsymbol{G}_{D_k}(\boldsymbol{G}_{D_k}^H(\boldsymbol{Q}_k^{-1} - \boldsymbol{T}_k\boldsymbol{T}_k^H)\boldsymbol{G}_{D_k})^{-1})$$
$$(66a)$$

$$= \sum_{k=1}^{K} \frac{\log\det(\widetilde{\boldsymbol{G}}_{D_k}^H\widetilde{\boldsymbol{G}}_{D_k})}{\log\det(\widetilde{\boldsymbol{G}}_{D_k}^H(\boldsymbol{I}_k^{-1} - \boldsymbol{Q}_k^{\frac{1}{2}}\boldsymbol{T}_k\boldsymbol{T}_k^H\boldsymbol{Q}_k^{\frac{1}{2}})\widetilde{\boldsymbol{G}}_{D_k})}, \quad (66b)$$

$$\overset{(a)}{\le} \sum_{k=1}^{K} \log\det\left(\big(\boldsymbol{I}_{N_{s_k}} - \boldsymbol{T}_k^H\boldsymbol{Q}_k\boldsymbol{T}_k\big)^{-1}\right) \quad (66c)$$

where the inequality (a) becomes equality when considering $\widetilde{\boldsymbol{G}}_{D_k} = \boldsymbol{Q}_k^{\frac{1}{2}}\boldsymbol{T}_k$ and thus the optimal $\boldsymbol{G}_{D_k} = \boldsymbol{Q}_k\boldsymbol{T}_k$ can be derived again, according to [43, Theorem 3]. Specifically, by applying [43, Theorem 3] to the function $\det(\widetilde{\boldsymbol{G}}_{D_k}^H(\boldsymbol{I}_{N_{s_k}}^{-1} - \boldsymbol{Q}_k^{\frac{1}{2}}\boldsymbol{T}_k\boldsymbol{T}_k^H\boldsymbol{Q}_k^{\frac{1}{2}})\widetilde{\boldsymbol{G}}_{D_k}) = \prod_{i=1}^{N_{s_k}} \sigma_i\big(\widetilde{\boldsymbol{G}}_{D_k}^H(\boldsymbol{I}_{N_{s_k}}^{-1} - \boldsymbol{Q}_k^{\frac{1}{2}}\boldsymbol{T}_k\boldsymbol{T}_k^H\boldsymbol{Q}_k^{\frac{1}{2}})\widetilde{\boldsymbol{G}}_{D_k})$, we readily infer that its minimization is achieved when the eigenspace of the optimal $\widetilde{\boldsymbol{G}}_{D_k}\widetilde{\boldsymbol{G}}_{D_k}^H$ matches with that of $\boldsymbol{Q}_k^{\frac{1}{2}}\boldsymbol{T}_k\boldsymbol{T}_k^H\boldsymbol{Q}_k^{\frac{1}{2}}$. Using this fact, it is further observed from (66b) that the singular values of $\widetilde{\boldsymbol{G}}_{D_k}$ actually have no influence on R_{sum} with a fractional structure, and thus we can simply set the optimal $\widetilde{\boldsymbol{G}}_{D_k}$ to $\widetilde{\boldsymbol{G}}_{D_k} = \boldsymbol{Q}_k^{\frac{1}{2}}\boldsymbol{T}_k$. It follows from (65a) and (66c) that the equivalence between the sum-rate maximization problem (4) and the WMMSE problem (6) can be established. This completes the proof.

APPENDIX B

Firstly, recalling $\boldsymbol{H}_{i,k} = \boldsymbol{U}_{i,k}\boldsymbol{\Lambda}_{i,k}\boldsymbol{V}_{i,k}^{\mathrm{H}}, \forall i,k$, we have

$$\lim_{N_{t_k}\to+\infty} \boldsymbol{U}_{i,k}\widetilde{\boldsymbol{\Lambda}}_{i,k}\boldsymbol{V}_{i,k}^{\mathrm{H}}\boldsymbol{V}_{k,k}\widetilde{\boldsymbol{\Lambda}}_{k,k}\boldsymbol{U}_{k,k}^{\mathrm{H}} = \boldsymbol{0}_{N_{r_i}\times N_{r_k}}, \forall i\ne k, \quad (67)$$

where $\widetilde{\boldsymbol{\Lambda}}_{i,k}=\frac{1}{\sqrt{N_{t_k}}}\boldsymbol{\Lambda}_{i,k}$. Since $\boldsymbol{U}_{i,k},\forall i,k$, are unitary, we have

$$\lim_{N_{t_k}\to+\infty}\widetilde{\boldsymbol{\Lambda}}_{i,k}\boldsymbol{V}_{i,k}^{\mathrm{H}}\boldsymbol{V}_{k,k}\widetilde{\boldsymbol{\Lambda}}_{k,k}=\mathbf{0}_{N_{r_i}\times N_{r_k}},\ \forall i\neq k. \quad (68)$$

Let $a_{p,q}$ be the (p,q)th element of the matrix $\boldsymbol{V}_{i,k}^{\mathrm{H}}(1:\mathrm{rank}(\boldsymbol{H}_{i,k}),:)\boldsymbol{V}_{k,k}(:,1:\mathrm{rank}(\boldsymbol{H}_{k,k}))$ with $p=1\cdots\mathrm{rank}(\boldsymbol{H}_{i,k})$ and $q=1\cdots\mathrm{rank}(\boldsymbol{H}_{k,k})$. Then, the (p,q)th element of the matrix $\widetilde{\boldsymbol{\Lambda}}_{i,k}\boldsymbol{V}_{i,k}^{\mathrm{H}}\boldsymbol{V}_{k,k}\widetilde{\boldsymbol{\Lambda}}_{k,k}$ can be expressed as $a_{p,q}\big[\widetilde{\boldsymbol{\Lambda}}_{i,k}\big]_{p,p}\big[\widetilde{\boldsymbol{\Lambda}}_{k,k}\big]_{q,q}$. Since the singular values $\big[\widetilde{\boldsymbol{\Lambda}}_{i,k}\big]_{l,l}$, $\forall i,k=1,\cdots,K$, are nonzero when $l\leq\mathrm{rank}(\boldsymbol{H}_{i,k})$, we readily conclude that the equality (68) holds if and only if $a_{p,q}=0$, $\forall p,q$, which leads to (25). This completes the proof.

APPENDIX C

Recalling the mmWave channel model (5), we have

$$\boldsymbol{H}_{k,i}=\sqrt{\frac{N_{r_k}N_{t_i}}{L_{k,i}}}\boldsymbol{A}_r^k\boldsymbol{\Lambda}_{k,i}\big(\boldsymbol{A}_t^i\big)^{\mathrm{H}},\ \forall k,i=1,\cdots,K, \quad (69)$$

where $\boldsymbol{A}_r^k=\big[\boldsymbol{a}_r(\theta_k^1),\cdots,\boldsymbol{a}_r(\theta_k^{L_{k,i}})\big]\in\mathbb{C}^{N_{r_k}\times L_{k,i}}$, $\boldsymbol{A}_t^i=\big[\boldsymbol{a}_t(\psi_i^1),\cdots,\boldsymbol{a}_t(\psi_i^{L_{k,i}})\big]\in\mathbb{C}^{N_{t_i}\times L_{k,i}}$ and $\boldsymbol{\Lambda}_{k,i}=\mathrm{diag}\big[\alpha_k^1,\cdots,\alpha_k^{L_{k,i}}\big]$. Note that $N_{s_k}\leq N_{t_k}^{RF}\leq L_{k,k}$ and $L_{k,i}=\mathrm{rank}(\boldsymbol{H}_{k,i}),\forall i,k$, are implied. Referring to [7], when $N_{t_k}\to+\infty$, the array steering vectors $\boldsymbol{a}_t^{\mathrm{H}}(\psi_i^l),\forall l$, are linearly independent and asymptotically orthogonal with probability one, i.e., $\lim_{N_{t_k}\to+\infty}\boldsymbol{a}_t^{\mathrm{H}}\big(\psi_i^{l_1}\big)\boldsymbol{a}_t\big(\psi_i^{l_2}\big)=0$, $\forall l_1\neq l_2$, and $\lim_{N_{t_k}\to+\infty}\big(\boldsymbol{A}_t^k\big)^{\mathrm{H}}\boldsymbol{A}_t^k=\boldsymbol{I}_{N_{t_k}}$, $\forall k$, which implies that in large-scale mmWave MIMO regime, the array response matrix $\boldsymbol{A}_t^k$ can be approximated to the right singular matrix of $\boldsymbol{H}_{i,k}$. Furthermore, by recalling (31) and exploiting the equality (25), the fully-digital BD-ZF precoder $\boldsymbol{F}_k^{\mathrm{ZF}}$ can be re-expressed as

$$\lim_{N_{t_k}\to+\infty}\boldsymbol{F}_k^{\mathrm{ZF}}=\boldsymbol{V}_{k,k}(:,1:L_{k,k})\sqrt{\overline{\boldsymbol{\Lambda}}_k}=\boldsymbol{A}_t^k\sqrt{\overline{\boldsymbol{\Lambda}}_k}, \quad (70)$$

where $\overline{\boldsymbol{\Lambda}}_k=\mathrm{BLKdiag}\big[\boldsymbol{\Lambda}_k,\mathbf{0}_{L_{k,k}-N_{s_k},L_{k,k}-N_{s_k}}\big]$ and $\boldsymbol{\Lambda}_k$ is determined by solving the problem (32). Obviously, the matrix $\boldsymbol{A}_t^k$ with unit-modulus elements can be realized by RF phase shifters, so that the proposed iterative-PP analog precoder $\boldsymbol{F}_{A_k}$ in (37) is easily obtained as $\boldsymbol{F}_{A_k}^\infty=\boldsymbol{A}_t^k(:,1:N_{t_k}^{RF})$ when $N_{t_k}\to+\infty$. Correspondingly, the optimal digital precoder is readily derived as $\boldsymbol{F}_{D_k}^\infty=\big[\sqrt{\boldsymbol{\Lambda}_k}\ \mathbf{0}_{N_{s_k}\times(N_{t_k}^{RF}-N_{s_k})}\big]^{\mathrm{H}}$. Using the above hybrid precoder design of the kth transmitter, the resultant interference at the ith receiver, where $i\neq k$, satisfies

$$\lim_{N_{t_k}\to+\infty}\boldsymbol{H}_{i,k}\boldsymbol{F}_{A_k}\boldsymbol{F}_{D_k}=\lim_{N_{t_k}\to+\infty}\boldsymbol{H}_{i,k}\boldsymbol{F}_{A_k}^\infty\boldsymbol{F}_{D_k}^\infty$$

$$=\lim_{N_{t_k}\to+\infty}\boldsymbol{U}_{i,k}\boldsymbol{\Lambda}_{i,k}\boldsymbol{V}_{i,k}^{\mathrm{H}}(1:L_{i,k},:)\boldsymbol{A}_t^k(:,1:N_{s_k})\sqrt{\boldsymbol{\Lambda}_k}, \quad (71)$$

$$=\lim_{N_{t_k}\to+\infty}\boldsymbol{U}_{i,k}\boldsymbol{\Lambda}_{i,k}\boldsymbol{V}_{i,k}^{\mathrm{H}}(1:L_{i,k},:)\boldsymbol{V}_{k,k}(:,1:N_{s_k})\sqrt{\boldsymbol{\Lambda}_k}=\mathbf{0}.$$

where the last equality holds by recalling (25). This completes the proof.

APPENDIX D

Firstly, let's define the objective function of the approximated WMMSE problem (8) as $f_{\mathrm{obj}}(\cdot)$, which can be rewritten as a quadratic function and a logarithmic function w.r.t each

variable of $\widetilde{\mathcal{A}}_k=\{\boldsymbol{G}_{A_k},\boldsymbol{G}_{D_k},\boldsymbol{F}_{A_k},\widetilde{\boldsymbol{F}}_{D_k}\}$ and $\boldsymbol{W}_k$, respectively, so we readily infer that all continuous partial derivatives of $f_{\mathrm{obj}}(\cdot)$ exist and thus $f_{\mathrm{obj}}(\cdot)$ is differentiable (continuous). Moreover, it is observed that the feasible region of problem (8) is the Cartesian product of five non-overlapped compact sets, i.e. $\mathcal{X}=\mathcal{X}_1\times\mathcal{X}_2\times\mathcal{X}_3\times\mathcal{X}_4\times\mathcal{X}_5$, which are respectively defined as

$$\mathcal{X}_1=\big\{\boldsymbol{X}\mid\boldsymbol{X}\in\mathbb{C}^{N_{r_k}^{RF}\times N_{s_k}},\forall k\big\},\mathcal{X}_2=\big\{\boldsymbol{X}\mid\boldsymbol{X}\in\mathbb{S}_{++}^{N_{s_k}},\forall k\big\},$$

$$\mathcal{X}_3=\big\{\boldsymbol{X}\mid\mathrm{Tr}(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X})\leq P_k,\forall k\big\},$$

$$\mathcal{X}_4=\big\{\boldsymbol{X}\mid|[\boldsymbol{X}]_{n,m}|=1,\boldsymbol{X}\in\mathbb{C}^{N_{t_k}\times N_{t_k}^{RF}},\forall k\big\},$$

$$\mathcal{X}_5=\big\{\boldsymbol{X}\mid|[\boldsymbol{X}]_{n,m}|=1,\boldsymbol{X}\in\mathbb{C}^{N_{r_k}\times N_{r_k}^{RF}},\forall k\big\} \quad (72)$$

As a result, in our proposed MM-Alt-Opt, the optimization variables of problem (8) can be decomposed into five independent blocks, namely, $\boldsymbol{G}_D\in\mathcal{X}_1$, $\boldsymbol{W}\in\mathcal{X}_2$, $\widetilde{\boldsymbol{F}}_D\in\mathcal{X}_3$, $\boldsymbol{F}_A\in\mathcal{X}_4$ and $\boldsymbol{G}_A\in\mathcal{X}_5$, $\forall k$. Notice that for simplicity, the notations $\{\boldsymbol{G}_{D_k},\forall k\}$, $\{\boldsymbol{W}_k,\forall k\}$, $\{\widetilde{\boldsymbol{F}}_{D_k}^{(I+1)},\forall k\}$, $\{\boldsymbol{F}_{A_k}^{(I+1)},\forall k\}$ and $\{\boldsymbol{G}_{A_k}^{(I+1)},\forall k\}$ are all shorted as $\boldsymbol{G}_D$, $\boldsymbol{W}$, $\widetilde{\boldsymbol{F}}_D$, $\boldsymbol{F}_A$ and $\boldsymbol{G}_A$, respectively. Then it follows from the general alternating optimization theory [36] that the proposed MM-Alt-Opt in essence belongs to the block successive upper-bound minimization method, of which the specific updates are summarized as (73) at the top of the next page.

According to the MM framework in Section IV-A, we find that the surrogate functions $g_{\mathrm{obj}}(\cdot)$ and $s_{\mathrm{obj}}(\cdot)$ (which are defined in (18) and (22), respectively, with their respective constant terms included) both are locally tight upper-bounds of the objective function $f_{\mathrm{obj}}(\cdot)$ in terms of different variables, and satisfy all conditions in Assumption 2 of [36].[1] Moreover, the common objective function $f_{\mathrm{obj}}(\cdot)$ in subproblems (73a)~(73c) can also be regarded as a globally tight upper-bound of itself. Following the MM theory and the philosiphy of alternating minimization, we now have

$$f_{\mathrm{obj}}\big(\{\boldsymbol{W}^{(I)},\boldsymbol{G}_A^{(I)},\boldsymbol{G}_D^{(I)},\boldsymbol{F}_A^{(I)},\widetilde{\boldsymbol{F}}_D^{(I)}\}\big)$$
$$\geq f_{\mathrm{obj}}\big(\{\boldsymbol{W}^{(I+1)},\boldsymbol{G}_A^{(I)},\boldsymbol{G}_D^{(I+1)},\boldsymbol{F}_A^{(I)},\widetilde{\boldsymbol{F}}_D^{(I+1)}\}\big)$$
$$=g_{\mathrm{obj}}\big(\boldsymbol{f}_A^{(I)};\boldsymbol{f}_A^{(I)}\big|\{\boldsymbol{W}^{(I+1)},\boldsymbol{G}_A^{(I)},\boldsymbol{G}_D^{(I+1)},\widetilde{\boldsymbol{F}}_D^{(I+1)}\}\big)$$
$$\geq g_{\mathrm{obj}}\big(\boldsymbol{f}_A^{(I+1)};\boldsymbol{f}_A^{(I)}\big|\{\boldsymbol{W}^{(I+1)},\boldsymbol{G}_A^{(I)},\boldsymbol{G}_D^{(I+1)},\widetilde{\boldsymbol{F}}_D^{(I+1)}\}\big)$$
$$=s_{\mathrm{obj}}\big(\boldsymbol{g}_A^{(I)};\boldsymbol{g}_A^{(I)}\big|\{\boldsymbol{W}^{(I+1)},\boldsymbol{G}_D^{(I+1)},\boldsymbol{F}_A^{(I+1)},\widetilde{\boldsymbol{F}}_D^{(I+1)}\}\big)$$
$$\geq s_{\mathrm{obj}}\big(\boldsymbol{g}_A^{(I+1)};\boldsymbol{g}_A^{(I)}\big|\{\boldsymbol{W}^{(I+1)},\boldsymbol{G}_D^{(I+1)},\boldsymbol{F}_A^{(I+1)},\widetilde{\boldsymbol{F}}_D^{(I+1)}\}\big)$$
$$\geq f_{\mathrm{obj}}\big(\{\boldsymbol{W}^{(I+1)},\boldsymbol{G}_A^{(I+1)},\boldsymbol{G}_D^{(I+1)},\boldsymbol{F}_A^{(I+1)},\widetilde{\boldsymbol{F}}_D^{(I+1)}\}\big) \quad (74)$$

Since the function $f_{\mathrm{obj}}(\cdot)$ in quadratic form is bounded below by zero, it is easily inferred that the sequence $\big\{f_{\mathrm{obj}}\big(\{\boldsymbol{W}^{(I)},\boldsymbol{G}_A^{(I)},\boldsymbol{G}_D^{(I)},\boldsymbol{F}_A^{(I)},\widetilde{\boldsymbol{F}}_D^{(I)}\}\big)\big\}$ obtained from (74) converges. Also, the continuity of $f_{\mathrm{obj}}(\cdot)$ implies

$$\lim_{I\to\infty}f_{\mathrm{obj}}\big(\{\boldsymbol{W}^{(I)},\boldsymbol{G}_A^{(I)},\boldsymbol{G}_D^{(I)},\boldsymbol{F}_A^{(I)},\widetilde{\boldsymbol{F}}_D^{(I)}\}\big)$$
$$=f_{\mathrm{obj}}\big(\{\boldsymbol{W}^{(\infty)},\boldsymbol{G}_A^{(\infty)},\boldsymbol{G}_D^{(\infty)},\boldsymbol{F}_A^{(\infty)},\widetilde{\boldsymbol{F}}_D^{(\infty)}\}\big). \quad (75)$$

---

[1] Notice that the overlapping essentially cyclic rule is actually adopted in our proposed MM-Alt-Opt, since both subproblems (73d) and (73e) are iteratively solved by the MM algorithm with the other variable blocks fixed. In this context, the convergence result remains the same as that based on the simple cyclic rule by referring to Corollary 2 in [36].

$$G_D^{(I+1)} = \arg \min_{G_D \in \mathcal{X}_1} f_{\text{obj}}\big(\{W^{(I)}, G_A^{(I)}, G_D, F_A^{(I)}, \widetilde{F}_D^{(I)}\}\big) \tag{73a}$$

$$W^{(I+1)} = \arg \min_{W \in \mathcal{X}_2} f_{\text{obj}}\big(\{W, G_A^{(I)}, G_D^{(I+1)}, F_A^{(I)}, \widetilde{F}_D^{(I)}\}\big) \tag{73b}$$

$$\widetilde{F}_D^{(I+1)} = \arg \min_{\widetilde{F}_D \in \mathcal{X}_3} f_{\text{obj}}\big(\{W^{(I+1)}, G_A^{(I)}, G_D^{(I+1)}, F_A^{(I)}, \widetilde{F}_D\}\big) \tag{73c}$$

$$F_A^{(I+1)} = \arg \min_{F_A \in \mathcal{X}_4} g_{\text{obj}}\big(f_A; f_A^{(I)} \big| \{W^{(I+1)}, G_A^{(I)}, G_D^{(I+1)}, \widetilde{F}_D^{(I+1)}\}\big) \tag{73d}$$

$$G_A^{(I+1)} = \arg \min_{G_A \in \mathcal{X}_5} s_{\text{obj}}\big(g_A; g_A^{(I)} \big| \{W^{(I+1)}, G_D^{(I+1)}, F_A^{(I+1)}, \widetilde{F}_D^{(I+1)}\}\big) \tag{73e}$$

To further validate that the proposed MM-Alt-Opt can converge to a (Boulingand) stationary point of problem (8), the required conditions by Theorem 2 in [36] are explored as follows. Firstly, it is clear that the functions $f_{\text{obj}}(\cdot)$, $g_{\text{obj}}(\cdot)$ and $s_{\text{obj}}(\cdot)$ all are convex quadratic functions. Secondly, the subproblems (73a) $\sim$ (73c) are all strictly convex problems with a unique optimal solution, while both subproblems (73d) and (73e) have unique solutions via phase projection. Based on the above two properties, we conclude that the whole sequence $\{W^{(I)}, G_A^{(I)}, G_D^{(I)}, F_A^{(I)}, \widetilde{F}_D^{(I)}\}$ generated by the proposed MM-Alt-Opt converges to the limit point $\{W^{(\infty)}, G_A^{(\infty)}, G_D^{(\infty)}, F_A^{(\infty)}, \widetilde{F}_D^{(\infty)}\}$, which is also a coordinatewise minimum point of problem (8). Furthermore, as a benefit of the differentiability of the objective function $f_{\text{obj}}(\cdot)$, this coordinatewise minimum point $\{W^{(\infty)}, G_A^{(\infty)}, G_D^{(\infty)}, F_A^{(\infty)}, \widetilde{F}_D^{(\infty)}\}$ is also a stationary point of problem (8). The interested readers may refer to [36] for the detailed proof.

Next, for the large-scale mmWave MIMO interference system, the tightness of the adopted approximation $F_{A_k}^H F_{A_k} \approx N_{t_k} I_{N_{t_K}^{RF}}$ is studied for demonstrating that the obtained $\{W^{(\infty)}, G_A^{(\infty)}, G_D^{(\infty)}, F_A^{(\infty)}, \widetilde{F}_D^{(\infty)}\}$ is also an asymptotically stationary point to the WMMSE problem (6). Specifically, based on Proposition 1, the orthogonality of the right singular subspaces of different transmit-receive channels further implies that the inter-user interference can be naturally eliminated with the aid of large-scale arrays. In this context, the K-user MIMO interference system can be decomposed into K independent parallel MIMO subsystems. Then according to the proposal of [7] for point-to-point MIMO systems, the optimal analog precoders of all K transmit-receiver pairs achieving the near-optimal sum rate performance can be obtained from a basis set of linearly independent array response vectors by exploiting the sparsity of mmWave channel. As a result, $F_{A_k}^H F_{A_k} \approx N_{t_k} I_{N_{t_K}^{RF}}$ is asymptotically satisfied when $N_{t_k} \to \infty$, since an inner product of two linearly independent array response vectors, i.e. $\sum_{n=0}^{N-1} e^{jan} = \frac{1 - e^{jaN}}{1 - e^{ja}}$ where a is the nonzero real number, and bounded compared to the infinite N_{t_k}.

REFERENCES

[1] L. Lu, *et al.*, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[2] F. Rusek, *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[3] Y. Wei, M. M. Zhao, M. J. Zhao and M. Lei, "Learned conjugate gradient descent network for massive MIMO detection," [Online]. Available: https://arxiv.org/abs/1906.03814.

[4] W. Roh, *et al.*, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.

[5] J. Hoydis, S. Ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[6] A. F. Molisch, *et al.*, "Hybrid beamforming for massive MIMO: A survey," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, Sep. 2017.

[7] O. El Ayach, *et al.*, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[8] G. M. Rebeiz, G. L. Tan, and J. S. Hayden, "RF MEMS phase shifters: Design and applications," *IEEE Microw. Mag.*, vol. 3, no. 2, pp. 72–81, Jun. 2002.

[9] R. Méndez Rial, *et al.*, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, Jan. 2016.

[10] Y. Zeng, R. Zhang, and Z. N. Chen, "Electromagnetic lens-focusing antenna enabled massive MIMO: Performance improvement and cost reduction," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1194–1206, Jun. 2014.

[11] 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.

[12] F. Sohrabi 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] F. Sohrabi and W. Yu, "Hybrid analog and digital beamforming for mmWave OFDM large-scale antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1432–1443, Jul. 2017.

[14] J. Singh and S. Ramakrishna, "On the feasibility of codebook-based beamforming in millimeter wave systems with multiple antenna arrays," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2670–2683, May 2015.

[15] W. Liu, *et al.*, "Partially-activated conjugate beamforming for LoS massive MIMO communications," *IEEE Access*, vol. 6, pp. 56504–56513, Oct. 2018.

[16] 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.

[17] Z. Zhou, N. Ge, Z. Wang, and S. Chen, "Hardware-efficient hybrid precoding for millimeter wave systems with multi-feed reflectarrays," *IEEE Access*, vol. 6, pp. 6795–6806, Mar. 2018.

[18] C. Xing, *et al.*, "A framework on hybrid MIMO transceiver design based on matrix-monotonic optimization," *IEEE Trans. Signal Process.*, vol. 67, no. 13, pp. 3531–3546, Jul. 2019.

[19] X. Yu, J. C. 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.

[20] W. Ni, X. Dong, and W. S. 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.

[21] C. E. 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.

[22] M. Á. Vázquez, L. Blanco and A. I. Pérez-Neira,"Hybrid AnalogDigital

Transmit Beamforming for Spectrum Sharing Backhaul Networks," *IEEE Trans. Signal Process.*, vol. 66, no. 9, pp. 2273-2285, May 2018.

[23] H. Ghauch, T. Kim, M. Bengtsson and M. Skoglund,"Subspace estimation and decomposition for large millimeter-wave MIMO systems" *IEEE J. Sel. Topics Signal Process.* vol. 10, no. 3, pp. 528–542, Apr. 2016.

[24] M. Kim and Y. H. Lee, "MSE-based hybrid RF/baseband processing for millimeter-wave communication systems in MIMO interference channels," *IEEE Trans. Veh. Techno.*, vol. 64, no. 6, pp. 2714–2720, Jun. 2015.

[25] D. H. Nguyen, L. B. Le, T. Le-Ngoc, and R. W. Heath, "Hybrid MMSE precoding and combining designs for mmWave multiuser systems," *IEEE Access*, vol. 5, pp. 19167–19181, Sep. 2017.

[26] 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.

[27] 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.

[28] X. Wu, D. Liu, and F. Yin, "Hybrid beamforming for multi-user massive MIMO systems," *IEEE Trans. Commun.*, vol. 66, no. 9, pp. 3879–3891, Sep. 2018.

[29] J. Zhang, M. Haardt, I. Soloeychik, and A. Wiesel,"A channel matching based hybrid analog-digital strategy for massive multi-user MIMO downlink systems," in *Proc. 9th IEEE SAM*, July 2016, pp. 1-5.

[30] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[31] M. Sadek, A. Tarighat, and A. H. Sayed, "A leakage-based precoding scheme for downlink multi-user MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1711–1721, May 2007.

[32] P. Cheng, M. Tao, and W. Zhang, "A new SLNR-based linear precoding for downlink multi-user multi-stream MIMO systems," *IEEE Commun. Lett.*, vol. 14, no. 11, pp. 1008–1010, Nov. 2010.

[33] A. Liu and V. Lau, "Phase only RF precoding for massive MIMO systems with limited RF chains," *IEEE Trans. Signal Process.*, vol. 62, no. 17, pp. 4505–4515, Sep. 2014.

[34] A. Liu and V. K. Lau, "Two-stage subspace constrained precoding in massive MIMO cellular systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3271–3279, Jun. 2015.

[35] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2017.

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

[37] D. P. Palomar, Y. Jiang, "MIMO transceiver design via majorization theory," *Foundations and Trends in Communications and Information Theory*, vol. 3, pp. 331–551, 2007.

[38] L. Wu, P. Babu, and D. P. Palomar, "Transmit waveform/receive filter design for MIMO radar with multiple waveform constraints," *IEEE Trans. Signal Process.*, vol. 66, no. 6, pp. 1526–1540, Mar. 2018.

[39] J. Pang, "Partially B-regular optimization and equilibrium problems," *Math. Oper. Res.*, vol. 32, no. 3, pp. 687–699, 2007.

[40] Q. Shi, M. Razaviyayn, Z. Q. 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.

[41] J. W. Wallace and M. A. Jensen, "Modeling the indoor MIMO wireless channel," *IEEE Trans. Antennas Propag.*, vol. 50, no. 5, pp. 591–599, May 2002.

[42] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*. Springer-Verlag: New York, 2011.

[43] B. Y. Wang, and B. Y. Xi, "Some inequalities for singular values of matrix products," *Linear algebra and its applications*, vol. 264, pp.109–115, 1997.