Alternating Optimization Based Hybrid Transceiver Designs for Wideband Millimeter-Wave Massive Multiuser MIMO-OFDM Systems

Minghao Yuan, Hua Wang, *Member, IEEE*, Hao Yin, and Dongxuan He

***Abstract*— Hybrid precoding has been considered as a promising technology for millimeter-wave (mmWave) massive multiple-input multiple-output (MIMO) systems, since it can achieve a tradeoff between system performance and hardware complexity. However, the optimal solution is difficult to obtain due to the coupling between analog precoder and digital precoder, as well as the non-convex constant modulus constraint, especially in multiuser scenarios. In this paper, we investigate several hybrid transceiver designs in wideband mmWave massive multiuser MIMO-OFDM systems for maximizing the spectral efficiency. Firstly, we propose two joint designs of hybrid precoder and combiner based on alternating optimization. Specifically, the intractable spectral efficiency maximization problem is reformulated as an equivalent weighted minimum mean square error (WMMSE) problem. To design the analog precoder and combiner with non-convex constant modulus constraint, we develop two efficient algorithms based on majorization minimization (MM) and element-wise block coordinate descent (EBCD) techniques, respectively. Secondly, to reduce the computational complexity, we propose a discrete Fourier transform (DFT) codebook based scheme, which can enhance the beamforming gain and mitigate the inter-beam interference. Thirdly, the convergence and complexity analysis are presented. The proposed two alternating optimization algorithms are guaranteed to converge to locally optimal solutions. Simulation results demonstrate that the proposed hybrid transceiver designs achieve significant performance gains over state-of-the-art schemes.**

***Index Terms*— Massive MIMO, millimeter-wave, multiuser MIMO, OFDM, hybrid transceiver designs, WMMSE, alternating optimization.**

I. Introduction

With the rapid development of various emerging applications such as virtual reality, industrial automation, and vehicle-to-everything, it is urgent to exploit higher frequency bands and increase spectral efficiency to meet the explosive demand for data rate and network capacity [1], [2]. Millimeter-wave (mmWave) bands are expected to provide enormous available bandwidth to alleviate the scarcity of spectrum resources [3], [4]. Moreover, the short wavelength at mmWave frequencies enables large-scale antenna arrays to be integrated in small physical dimension. Massive multiple-input multiple-output (MIMO) can provide sufficient beamforming gain to overcome the severe path loss of mmWave signals. By using precoding technique, massive MIMO is capable of achieving spatial multiplexing gain to enhance spectral efficiency significantly [5], [6]. Therefore, mmWave massive MIMO has been widely recognized as one of the key enabling technologies in fifth generation (5G) and future wireless communication systems [7], [8].

However, conventional fully digital precoding requires one dedicated radio frequency (RF) chain for each antenna element, which is impractical in mmWave massive MIMO systems due to prohibitive hardware cost and power consumption. To address this challenge, hybrid analog and digital precoding has been proposed in [9], which is implemented using high-dimensional analog precoder for enabling beamforming gain, followed by low-dimensional digital precoder for obtaining spatial multiplexing gain. Compared with fully digital precoding, hybrid precoding can significantly reduce the number of RF chains without excessive performance degradation, thereby achieving a tradeoff between system performance and hardware complexity [10], [11], [12].

A. Previous Works and Motivation

In recent years, hybrid precoding designs for mmWave massive MIMO systems have received considerable interest from both academia and industry. Several pioneering works for narrowband single-user MIMO (SU-MIMO) systems were proposed in [9], [13], and [14]. The main idea is to first decouple the transmitter and receiver, and then jointly design the analog precoder and digital precoder to approximate the fully digital solution. Specifically, it was shown in [9] and [13] that maximizing the spectral efficiency can be approximated by minimizing the Euclidean distance between the hybrid precoder and the optimal fully digital precoder. By exploiting the sparse property of mmWave channels, the hybrid precoding design was reformulated as a sparse

Manuscript received 30 August 2022; revised 25 January 2023; accepted 7 April 2023. Date of publication 27 April 2023; date of current version 12 December 2023. This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFB1807900; in part by the National Natural Science Foundation of China under Grant 62101306; in part by the State Key Laboratory of Wireless Mobile Communications, China Academy of Telecommunications Technology (CATT); and in part by Datang Linktester Technology Company Ltd. The associate editor coordinating the review of this article and approving it for publication was Y. Wu. *(Corresponding authors: Hua Wang; Dongxuan He.)*

Minghao Yuan, Hua Wang, and Dongxuan He are with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: minghaoyuan@bit.edu.cn; hwang139@sohu.com; dongxuan_he@bit.edu.cn).

Hao Yin is with the Institute of China Electronic System Engineering Corporation, Beijing 100141, China (e-mail: yinhao@cashq.ac.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2023.3269056.

Digital Object Identifier 10.1109/TWC.2023.3269056

reconstruction problem and solved by the orthogonal matching pursuit (OMP) algorithm [9]. To deal with the non-convex constant modulus constraint imposed by the analog precoder, a manifold optimization based alternating minimization (MO-AltMin) algorithm was proposed in [13], which performed close to the optimal fully digital precoder at the expense of high computational complexity. In addition, it was proved in [14] that the hybrid precoding architecture can realize the same performance as any fully digital precoder when the number of RF chains is twice the total number of data streams, and an iterative coordinate descent algorithm was developed to achieve near-optimal performance with fewer number of RF chains.

Moreover, there are various works on hybrid precoding design for narrowband multiuser MIMO (MU-MIMO) systems [15], [16], [17], where the two-stage approach was widely adopted. In [15], a two-stage hybrid precoding scheme was proposed for downlink multiuser mmWave systems. In the first stage, the analog precoder at the base station (BS) and the analog combiners of all users were jointly designed to maximize the beamforming gain. In the second stage, the zero-forcing (ZF) based digital precoder was employed to cancel the multiuser interference. Nonetheless, this work was limited to only single-stream transmission for each user. To support multi-stream transmission per user for improving spatial multiplexing gain, a hybrid block diagonalization (BD) scheme was proposed in [16], where the analog precoder was first designed by the equal gain transmission (EGT) method to harvest the large array gain, then the BD technique was performed in the baseband to eliminate the multiuser interference. Besides, to further enhance both spectral efficiency and bit error rate (BER) performance, a nonlinear hybrid transceiver design based on the Gram-Schmidt orthogonalization (GSO) and block diagonal uniform channel decomposition (BD-UCD) was proposed in [17].

The aforementioned works [9], [13], [14], [15], [16], [17] mainly focus on narrowband transmission under flat fading channels. To take full advantage of the large available bandwidth in mmWave bands to achieve gigabits-per-second (Gbps) data rate, orthogonal frequency division multiplexing (OFDM) based hybrid precoding designs under wideband frequency-selective channels have recently attracted tremendous attention [18], [19], [20], [21], [22]. Nevertheless, the approaches designed for narrowband mmWave systems cannot be directly extended to wideband mmWave systems. This is due to the fact that for OFDM based wideband systems, a common analog precoder is shared by all subcarriers while different digital precoders are designed for different subcarriers, which makes hybrid precoding designs in wideband scenarios more complicated and challenging.

Hybrid precoding designs for wideband SU-MIMO-OFDM systems have been extensively investigated in [18], [19], and [20]. To be specific, it was shown in [18] that the channel covariance matrices of different subcarriers are approximately the same when the number of antennas is sufficiently large, and a heuristic algorithm was proposed for wideband mmWave systems by exploiting the average of the channel covariance matrices over all subcarriers. To optimize the transmission reliability, a manifold optimization (MO) based algorithm under the minimum mean square error (MMSE) criterion was proposed in [19]. Additionally, a principal component analysis (PCA) based wideband hybrid precoding approach was proposed in [20], where the frequency-flat analog precoder was extracted from the frequency-selective fully digital precoder. Unfortunately, the above works [18], [19], [20] are restricted to SU-MIMO-OFDM systems, which cannot be directly extended to MU-MIMO-OFDM systems due to the existence of multiuser interference.

Recently, a few hybrid precoding schemes for wideband MU-MIMO-OFDM systems have been studied in [21] and [22]. Specifically, an alternating minimization algorithm by phase extraction (AM-PE) was proposed to approximate the fully digital precoder using the analog precoder in [21]. In [22], a two-stage hybrid transceiver design was proposed for downlink MU-MIMO-OFDM systems, where the analog precoder and combiner design was first formulated as a constrained Tucker2 tensor decomposition (TD) problem and solved by the alternating least square algorithm, then the digital precoder and combiner were obtained by the regularized channel diagonalization (RCD) method to balance multiuser interference and noise suppression. However, this approach is limited to the case where the number of RF chains is exactly equal to that of data streams.

Although hybrid transceiver designs for mmWave massive MIMO systems have been extensively investigated, the optimal solution in the sense of spectral efficiency is still unknown [13], [23]. To the best of our knowledge, instead of directly solving the spectral efficiency maximization problem, most of existing works indirectly optimize the spectral efficiency by the matrix factorization [9], [13], [20], [21] or the two-stage processing [15], [16], [17], [22], which inevitably result in certain performance degradation compared to the optimal solution. Moreover, it is well known that dirty paper coding (DPC) [24] achieves the sum capacity of downlink MU-MIMO systems [25], [26], [27], [28], which can serve as an upper bound of achievable spectral efficiency.[1] Nevertheless, there is still a significant gap between the spectral efficiency of state-of-the-art hybrid transceiver design and the sum capacity [14], [23], especially in MU-MIMO-OFDM systems. The above facts motivate us to consider more efficient solutions to reduce the gap with sum capacity.

B. Contributions

In this paper, we investigate hybrid transceiver designs for wideband mmWave massive MU-MIMO-OFDM systems. In particular, compared to the existing works [9], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], a more general scenario with multiple antennas, multiple RF chains, and multiple data streams at both the transmitter and receivers is

[1]Note that spectral efficiency refers to achievable rate in the single-user case and sum rate in the multiuser case throughout this paper. Sum capacity (bps/Hz) refers to the maximum sum rate in a MU-MIMO system [25], [26], [27], [28].

considered. The main contributions of this paper are summarized as follows:

- Firstly, to maximize the spectral efficiency, we propose two joint hybrid transceiver designs based on alternating optimization. Different from the matrix factorization for approximating fully digital solution [9], [13], [20], [21] and two-stage processing for suppressing multiuser interference [15], [16], [17], [22] in most of the existing works, we tackle the original spectral efficiency maximization problem directly. However, this optimization problem is non-convex and NP-hard due to the highly coupled variables and the constant modulus constraint imposed on the analog precoder and combiner. By exploiting the equivalence between the spectral efficiency maximization problem and the weighted minimum mean square error (WMMSE) problem [30], [31], we transform the complicated problem into a tractable equivalent one. Based on alternating optimization framework, the WMMSE problem is decoupled into a series of sub-problems with respect to each variable, which are solved separately.
- Secondly, to deal with the sub-problems of the analog precoder and combiner with non-convex constant modulus constraint, we propose a majorization minimization (MM) based algorithm and an element-wise block coordinate descent (EBCD) based algorithm, respectively. Moreover, the closed-form solutions of the digital precoder and combiner are obtained by utilizing the Lagrangian multiplier method and the first-order optimality condition, respectively. Each variable is alternately updated until convergence.
- Thirdly, to reduce the computational complexity, we propose a discrete Fourier transform (DFT) codebook based two-stage hybrid transceiver design. In the first stage, the analog precoder and combiner are jointly designed by searching from the DFT codebook, which effectively enhances the beamforming gain and combats the inter-beam interference. In the second stage, the BD technique is employed to further eliminate the multiuser interference. Furthermore, the proposed DFT-BD scheme can be utilized for the initialization of the above two alternating optimization algorithms to speed up the convergence.
- Finally, the convergence analysis, complexity analysis, and numerical results are presented to verify the effectiveness of the proposed schemes. The convergence analysis reveals that the proposed two alternating optimization algorithms based on MM and EBCD are guaranteed to converge to locally optimal solutions. The simulation results demonstrate that the proposed alternating optimization based hybrid transceiver designs achieve significant performance gains over existing schemes in terms of spectral efficiency and bit error rate (BER) under various conditions. Also, the proposed DFT-BD scheme achieves comparable performance to state-of-the-art benchmark with reduced computational complexity.

The rest of this paper is organized as follows. Section II introduces the system model and problem formulation. In Section III, we reformulate the spectral efficiency maximization problem as an equivalent WMMSE problem and propose two joint hybrid transceiver designs based on alternating optimization. Section IV proposes a low-complexity DFT codebook based two-stage hybrid transceiver design. The convergence analysis and complexity analysis are presented in Section V. In Section VI, we provide extensive simulation results to verify the effectiveness of the proposed schemes. Finally, we conclude this paper in Section VII.

The notations used throughout this paper are illustrated as follows. Boldface lower-case letters indicate column vectors and boldface upper-case letters indicate matrices, respectively. $\mathbb{C}$ denotes the set of complex numbers. $(\cdot)^*$, $(\cdot)^T$, and $(\cdot)^H$ denote the operation of conjugate, transpose, and conjugate transpose of a vector or a matrix, respectively. $|\cdot|$, $(\cdot)^{-1}$, $\|\cdot\|_F$, $\mathrm{Tr}(\cdot)$, and $\mathrm{rank}(\cdot)$ represent the determinant (or module of a complex number), inversion, Frobenius norm, trace, and rank of a matrix, respectively. $[\mathbf{A}]_{i,:}$, $[\mathbf{A}]_{:,j}$, and $[\mathbf{A}]_{i,j}$ represent the i-th row, the j-th column, and the (i,j)-th element of matrix $\mathbf{A}$, respectively. $\mathrm{blkdiag}\{\mathbf{A}_1,\ldots,\mathbf{A}_N\}$ represents the block diagonal matrix with diagonal elements $\mathbf{A}_1,\ldots,\mathbf{A}_N$. $\mathrm{vec}(\mathbf{A})$ denotes the vectorization of $\mathbf{A}$ and $\mathrm{unvec}_{M,N}(\mathbf{a})$ denotes the transformation from the $MN\times 1$ vector $\mathbf{a}$ to an $M\times N$ matrix. $\lambda_{max}(\mathbf{A})$ represents the maximum eigenvalue of $\mathbf{A}$. $\arg(\cdot)$ denotes the phase extraction. $\otimes$ denotes the Kronecker product. $\mathbf{I}_N$ represents the $N\times N$ identity matrix. $\mathbf{x}\sim\mathcal{CN}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ denotes that $\mathbf{x}$ is a circularly symmetric complex Gaussian (CSCG) vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. $\mathbb{E}\{\cdot\}$ represents the mathematical expectation. $\Re\{\cdot\}$ denotes the real part of a complex number. The major variables adopted in the paper are listed in Table I for ease of reference.

II. System Model and Problem Formulation

A. System Model

As shown in Fig. 1, we consider the downlink transmission of a wideband mmWave massive MU-MIMO-OFDM system with K subcarriers, where the fully-connected hybrid architecture is adopted at the base station (BS) and all users. Specifically, the BS equipped with N_t transmit antennas and N_t^{RF} RF chains communicates with U users, and each user is equipped with N_r receive antennas and N_r^{RF} RF chains, the BS transmits N_s data streams to each user per subcarrier. To ensure the effectiveness of transmission, the numbers of RF chains at the BS and each user should satisfy $UN_s\le N_t^{\mathrm{RF}}\ll N_t$ and $N_s\le N_r^{\mathrm{RF}}\ll N_r$, respectively.

At the BS, the transmitted symbol vector $\mathbf{s}[k]=[\mathbf{s}_1^T[k],\ldots,\mathbf{s}_U^T[k]]^T\in\mathbb{C}^{UN_s\times 1}$ for all users at the k-th subcarrier is first precoded by the digital precoder $\mathbf{F}_{\mathrm{BB}}[k]=[\mathbf{F}_{\mathrm{BB},1}[k],\ldots,\mathbf{F}_{\mathrm{BB},U}[k]]\in\mathbb{C}^{N_t^{\mathrm{RF}}\times UN_s}$, where $\mathbf{s}_u[k]\in\mathbb{C}^{N_s\times 1}$ and $\mathbf{F}_{\mathrm{BB},u}[k]\in\mathbb{C}^{N_t^{\mathrm{RF}}\times N_s}$ represent the symbol vector and digital precoder for the u-th user at the k-th subcarrier, respectively, and $\mathbb{E}\left\{\mathbf{s}[k]\mathbf{s}^H[k]\right\}=\mathbf{I}_{UN_s}$. Then, the precoded signal is transformed into the time domain by using K-point inverse fast Fourier transform (IFFT), and a sufficiently long cyclic prefix (CP) is added to avoid inter-symbol interference. After passing through the RF chains, the transmitted signal is further precoded by the analog precoder $\mathbf{F}_{\mathrm{RF}}\in\mathbb{C}^{N_t\times N_t^{\mathrm{RF}}}$.

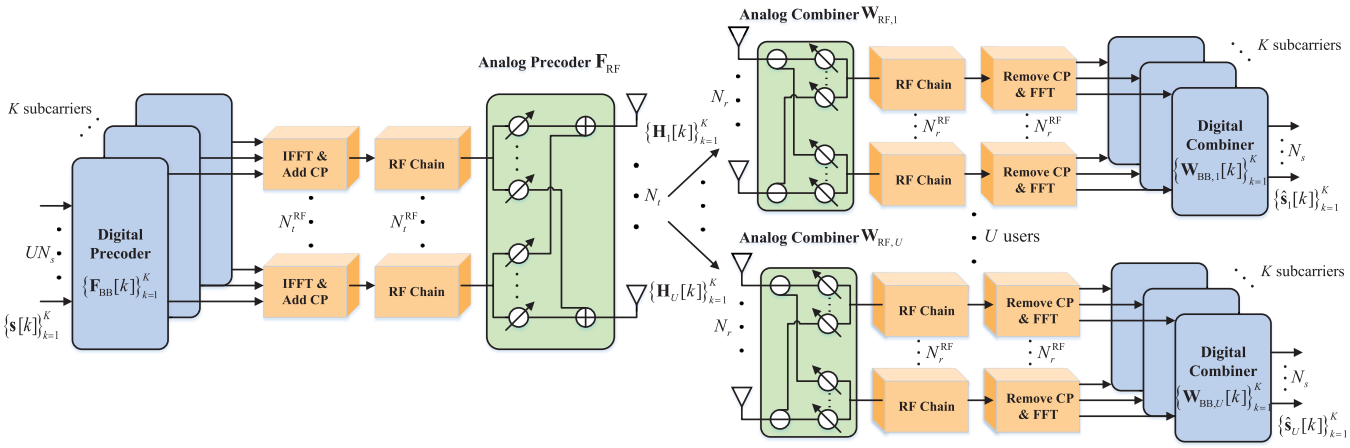

Fig. 1. Block diagram of hybrid transceiver for a mmWave massive MU-MIMO-OFDM system.

TABLE I
VARIABLE LIST

Notation	Definition
$\mathbf{F}_{\mathrm{RF}}$,$\mathbf{F}_{\mathrm{BB}}[k]$	Analog precoder and digital precoder
$\mathbf{W}_{\mathrm{RF},u}$,$\mathbf{W}_{\mathrm{BB},u}[k]$	Analog combiner and digital combiner
$\tilde{\mathbf{F}}_{\mathrm{RF}}$,$\tilde{\mathbf{F}}_{\mathrm{BB}}[k]$	Effective analog precoder and digital precoder
$\mathbf{\Lambda}_u[k]$	Weight matrix
$\mathbf{H}_u[k]$	Channel matrix
$\mathbf{s}_u[k]$	Transmitted signal
$\hat{\mathbf{s}}_u[k]$	Received signal
$\mathbf{n}_u[k]$	Noise
N_t,N_r	Numbers of antennas at the BS and each user
N_t^{RF},N_r^{RF}	Numbers of RF chains at the BS and each user
U	Number of users
K	Number of subcarriers
N_s	Number of data streams per user per subcarrier
P_k	Maximum transmit power per subcarrier
σ_n^2	Noise variance
L_u	Number of paths
$\alpha_{u,l}$,$\tau_{u,l}$	Complex gain and delay
$\theta_{u,l}$,$\phi_{u,l}$	AoA and AoD
$\mathbf{a}_r(\theta_{u,l})$,$\mathbf{a}_t(\phi_{u,l})$	Array response vectors of Rx and Tx

It is worth noting that the analog precoder is common for all subcarriers since it operates after IFFT. In other words, the analog precoder is frequency-flat whereas the digital precoder is frequency-selective, which is the main difference between wideband mmWave systems and narrowband mmWave systems.

At the u-th user, the received signal is first processed by the analog combiner $\mathbf{W}_{\mathrm{RF},u} \in \mathbb{C}^{N_r \times N_r^{\mathrm{RF}}}$, which is also shared by all subcarriers. After removing the CP and implementing K-point fast Fourier transform (FFT), the frequency domain signal at each subcarrier is further processed through the digital combiner $\mathbf{W}_{\mathrm{BB},u}[k] \in \mathbb{C}^{N_r^{\mathrm{RF}} \times N_s}$ for $k = 1, \ldots, K$. Finally, the combined signal of the u-th user at the k-th subcarrier can be expressed as

$$\begin{aligned}\hat{\mathbf{s}}_u[k] =& \mathbf{W}_{\mathrm{BB},u}^H[k]\mathbf{W}_{\mathrm{RF},u}^H\mathbf{H}_u[k]\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}[k]\mathbf{s}[k] \\ &+ \mathbf{W}_{\mathrm{BB},u}^H[k]\mathbf{W}_{\mathrm{RF},u}^H\mathbf{n}_u[k] \\ =& \mathbf{W}_{\mathrm{BB},u}^H[k]\mathbf{W}_{\mathrm{RF},u}^H\mathbf{H}_u[k]\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB},u}[k]\mathbf{s}_u[k] \\ &+ \sum_{i\neq u}\mathbf{W}_{\mathrm{BB},u}^H[k]\mathbf{W}_{\mathrm{RF},u}^H\mathbf{H}_u[k]\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB},i}[k]\mathbf{s}_i[k] \\ &+ \mathbf{W}_{\mathrm{BB},u}^H[k]\mathbf{W}_{\mathrm{RF},u}^H\mathbf{n}_u[k], \end{aligned} \tag{1}$$

where $\mathbf{H}_u[k] \in \mathbb{C}^{N_r \times N_t}$ represents the frequency domain channel between the BS and the u-th user at the k-th subcarrier, $\mathbf{n}_u[k] \in \mathbb{C}^{N_r \times 1}$ denotes the additive white Gaussian noise (AWGN) obeying $\mathcal{CN}(\mathbf{0}, \sigma_n^2\mathbf{I}_{N_r})$. Note that the analog precoder at the BS and the analog combiners for all users are implemented using phase shifters, where each element should satisfy the constant modulus constraint, i.e., $\left|[\mathbf{F}_{\mathrm{RF}}]_{i,j}\right| = \frac{1}{\sqrt{N_t}}$, $\forall i, j$ and $\left|[\mathbf{W}_{\mathrm{RF},u}]_{m,n}\right| = \frac{1}{\sqrt{N_r}}$, $\forall u, m, n$. Moreover, the transmit power constraint at the k-th subcarrier is given by $\|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}[k]\|_F^2 \leq P_k$ for $k = 1, \ldots, K$.

B. Channel Model

In this paper, we adopt the geometric channel model [21] to characterize the limited scattering nature of wideband mmWave channels. It is assumed that uniform linear arrays (ULA) are deployed at the BS and all users. According to [21], the delay-d channel matrix between the BS and the u-th user can be expressed as

$$\mathbf{H}_{u,d} = \sqrt{\frac{N_r N_t}{L_u}} \sum_{l=1}^{L_u} \alpha_{u,l} p(dT_s - \tau_{u,l}) \mathbf{a}_r(\theta_{u,l}) \mathbf{a}_t^H(\phi_{u,l}), \tag{2}$$

where L_u represents the number of propagation paths between the BS and the u-th user, $\alpha_{u,l} \sim \mathcal{CN}(0, 1)$ and $\tau_{u,l}$ denote the complex gain and delay of the l-th path, respectively, $p(\tau)$ is the pulse shaping response sampled at τ, T_s is the

sampling interval, $\theta_{u,l}$ and $\phi_{u,l}$ represent the angle of arrival (AoA) and angle of departure (AoD) of the l-th path, respectively. With half-wavelength antenna spacing, the transmit and receive array response vectors $\mathbf{a}_t(\phi_{u,l})$ and $\mathbf{a}_r(\theta_{u,l})$ can be respectively represented as

$$\mathbf{a}_t\left(\phi_{u,l}\right) = \frac{1}{\sqrt{N_t}}\left[1, e^{j\pi\sin(\phi_{u,l})}, \ldots, e^{j(N_t-1)\pi\sin(\phi_{u,l})}\right]^T, \tag{3a}$$

$$\mathbf{a}_r\left(\theta_{u,l}\right) = \frac{1}{\sqrt{N_r}}\left[1, e^{j\pi\sin(\theta_{u,l})}, \ldots, e^{j(N_r-1)\pi\sin(\theta_{u,l})}\right]^T. \tag{3b}$$

Furthermore, the frequency domain channel matrix for the u-th user at the k-th subcarrier can be expressed as

$$\mathbf{H}_u[k] = \sum_{d=0}^{D-1} \mathbf{H}_{u,d} \mathrm{e}^{-j2\pi\frac{k}{K}d}. \tag{4}$$

Also, the frequency domain channel $\mathbf{H}_u[k]$ can be rewritten as an equivalent form, given by

$$\mathbf{H}_u[k] = \mathbf{\Delta}_{r,u}\mathbf{D}_u[k]\mathbf{\Delta}_{t,u}^H, \tag{5}$$

where $\mathbf{D}_u[k] \in \mathbb{C}^{L_u \times L_u}$ is a diagonal matrix, $\mathbf{\Delta}_{r,u} \in \mathbb{C}^{N_r \times L_u}$ and $\mathbf{\Delta}_{t,u} \in \mathbb{C}^{N_t \times L_u}$ can be respectively written as

$$\mathbf{\Delta}_{r,u} = \left[\mathbf{a}_r(\theta_{u,1}), \ldots, \mathbf{a}_r(\theta_{u,L_u})\right], \tag{6a}$$

$$\mathbf{\Delta}_{t,u} = \left[\mathbf{a}_t(\phi_{u,1}), \ldots, \mathbf{a}_r(\phi_{u,L_u})\right]. \tag{6b}$$

C. Problem Formulation

In this paper, we investigate hybrid transceiver designs for wideband mmWave massive MU-MIMO-OFDM systems. Assuming that Gaussian symbols are transmitted, the achievable spectral efficiency of the u-th user at the k-th subcarrier can be expressed as

$$R_{u,k} = \log\left|\mathbf{I}_{N_s} + (\mathbf{X}_u[k])^{-1}\mathbf{W}_{\mathrm{BB},u}^H[k]\mathbf{W}_{\mathrm{RF},u}^H\mathbf{H}_u[k]\mathbf{F}_{\mathrm{RF}} \times \mathbf{F}_{\mathrm{BB},u}[k]\mathbf{F}_{\mathrm{BB},u}^H[k]\mathbf{F}_{\mathrm{RF}}^H\mathbf{H}_u^H[k]\mathbf{W}_{\mathrm{RF},u}\mathbf{W}_{\mathrm{BB},u}[k]\right|, \tag{7}$$

where $\mathbf{X}_u[k]$ is the covariance matrix of multiuser interference plus noise, given by

$$\mathbf{X}_u[k] = \sum_{i\neq u}\Big(\mathbf{W}_{\mathrm{BB},u}^H[k]\mathbf{W}_{\mathrm{RF},u}^H\mathbf{H}_u[k]\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB},i}[k] \times \mathbf{F}_{\mathrm{BB},i}^H[k]\mathbf{F}_{\mathrm{RF}}^H\mathbf{H}_u^H[k]\mathbf{W}_{\mathrm{RF},u}\mathbf{W}_{\mathrm{BB},u}[k]\Big) + \sigma_n^2\mathbf{W}_{\mathrm{BB},u}^H[k]\mathbf{W}_{\mathrm{RF},u}^H\mathbf{W}_{\mathrm{RF},u}\mathbf{W}_{\mathrm{BB},u}[k]. \tag{8}$$

Furthermore, we assume that perfect channel state information (CSI) is available at the transmitter and all receivers. In practice, CSI can be efficiently obtained by the state-of-the-art channel estimation techniques [32], [33]. In this paper, our objective is to jointly design the analog precoder $\mathbf{F}_{\mathrm{RF}}$ and the digital precoder $\{\mathbf{F}_{\mathrm{BB}}[k]\}_{k=1}^K$ at the BS, as well as the analog combiner $\{\mathbf{W}_{\mathrm{RF},u}\}_{u=1}^U$ and the digital combiner $\{\{\mathbf{W}_{\mathrm{BB},u}[k]\}_{u=1}^U\}_{k=1}^K$ at all users for maximizing the average spectral efficiency over all subcarriers, subject to the transmit power constraint per subcarrier and the constant modulus constraint imposed by $\mathbf{F}_{\mathrm{RF}}$ and $\{\mathbf{W}_{\mathrm{RF},u}\}_{u=1}^U$. Specifically, the spectral efficiency maximization problem can be formulated as follows

$$\begin{aligned}\max_{\substack{\mathbf{F}_{\mathrm{RF}},\mathbf{F}_{\mathrm{BB}}[k],\\ \mathbf{W}_{\mathrm{RF},u},\mathbf{W}_{\mathrm{BB},u}[k]}} \quad & \frac{1}{K}\sum_{k=1}^K\sum_{u=1}^U R_{u,k} \\ \text{s.t.} \quad & \|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}[k]\|_F^2 \le P_k, \quad \forall k, \\ & \left|[\mathbf{F}_{\mathrm{RF}}]_{i,j}\right| = \frac{1}{\sqrt{N_t}}, \quad \forall i,j, \\ & \left|[\mathbf{W}_{\mathrm{RF},u}]_{m,n}\right| = \frac{1}{\sqrt{N_r}}, \quad \forall u,m,n. \end{aligned} \tag{9}$$

Nevertheless, we observe that problem (9) is non-convex and NP-hard due to the sophisticated objective function, the highly coupled optimization variables, and the non-convex constant modulus constraint. Consequently, obtaining the globally optimal solution of problem (9) is extremely challenging, whereas it is more common to acquire its locally optimal solution. In fact, compared with the theoretical upper bound of spectral efficiency achieved by DPC, all the existing approaches such as the matrix factorization for approximating fully digital solution [20], [21] and the two-stage processing for suppressing multiuser interference [22] suffer from non-negligible performance deterioration. To address this challenge, we adopt alternating optimization framework for hybrid transceiver design and propose several efficient solutions in the next section.

III. Alternating Optimization Based Hybrid Transceiver Designs

In this section, the spectral efficiency maximization problem is first reformulated as a tractable WMMSE problem in virtue of their equivalence. Then, we propose two alternating optimization based hybrid transceiver designs by applying the majorization minimization (MM) method and the element-wise block coordinate descent (EBCD) method.

A. Equivalent WMMSE Problem

In this subsection, we establish the equivalence between the spectral efficiency maximization problem and the WMMSE problem. In addition, the optimal closed-form solutions of the digital combiner and auxiliary weight matrix are given.

Mean square error (MSE) is a typical performance metric to characterize the transmission reliability [19]. Specifically, the MSE matrix between the received signal and the transmitted signal for the u-th user at the k-th subcarrier can be defined as

$$\begin{aligned}\mathbf{E}_u[k] &\triangleq \mathbb{E}\left\{(\hat{\mathbf{s}}_u[k]-\mathbf{s}_u[k])(\hat{\mathbf{s}}_u[k]-\mathbf{s}_u[k])^H\right\} \\ &= \mathbf{W}_{\mathrm{BB},u}^H[k]\mathbf{W}_{\mathrm{RF},u}^H\mathbf{H}_u[k]\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB},u}[k] \\ &\quad\times \mathbf{F}_{\mathrm{BB},u}^H[k]\mathbf{F}_{\mathrm{RF}}^H\mathbf{H}_u^H[k]\mathbf{W}_{\mathrm{RF},u}\mathbf{W}_{\mathrm{BB},u}[k] \\ &\quad+\sum_{i\neq u}\Big(\mathbf{W}_{\mathrm{BB},u}^H[k]\mathbf{W}_{\mathrm{RF},u}^H\mathbf{H}_u[k]\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB},i}[k] \\ &\quad\times \mathbf{F}_{\mathrm{BB},i}^H[k]\mathbf{F}_{\mathrm{RF}}^H\mathbf{H}_u^H[k]\mathbf{W}_{\mathrm{RF},u}\mathbf{W}_{\mathrm{BB},u}[k]\Big) \\ &\quad-2\Re\left\{\mathbf{W}_{\mathrm{BB},u}^H[k]\mathbf{W}_{\mathrm{RF},u}^H\mathbf{H}_u[k]\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB},u}[k]\right\} \\ &\quad+\sigma_n^2\mathbf{W}_{\mathrm{BB},u}^H[k]\mathbf{W}_{\mathrm{RF},u}^H\mathbf{W}_{\mathrm{RF},u}\mathbf{W}_{\mathrm{BB},u}[k]+\mathbf{I}_{N_s}. \end{aligned} \tag{10}$$

By introducing the auxiliary weight matrix $\mathbf{\Lambda}_u[k] \in \mathbb{C}^{N_s \times N_s}$, for $u = 1, \ldots, U, k = 1, \ldots, K$, the spectral efficiency maximization problem (9) can be reformulated as the following WMMSE problem

$$\begin{aligned} \min_{\substack{\mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{BB}}[k], \\ \mathbf{W}_{\mathrm{RF},u}, \mathbf{W}_{\mathrm{BB},u}[k], \mathbf{\Lambda}_u[k]}} \quad & \frac{1}{K} \sum_{k=1}^{K} \sum_{u=1}^{U} \Big(\mathrm{Tr}(\mathbf{\Lambda}_u[k] \mathbf{E}_u[k]) \\ & - \log |\mathbf{\Lambda}_u[k]| - N_s \Big) \\ \text{s.t.} \quad & \|\mathbf{F}_{\mathrm{RF}} \mathbf{F}_{\mathrm{BB}}[k]\|_F^2 \le P_k, \quad \forall k, \\ & \left| [\mathbf{F}_{\mathrm{RF}}]_{i,j} \right| = \frac{1}{\sqrt{N_t}}, \quad \forall i, j, \\ & \left| [\mathbf{W}_{\mathrm{RF},u}]_{m,n} \right| = \frac{1}{\sqrt{N_r}}, \quad \forall u, m, n, \end{aligned} \tag{11}$$

where $\mathbf{\Lambda}_u[k] \succeq \mathbf{0}$ is a positive semi-definite matrix. The following proposition establishes the equivalence between the spectral efficiency maximization problem and the WMMSE problem.

Proposition 1: The WMMSE problem (11) is equivalent to the spectral efficiency maximization problem (9), in the sense that the globally optimal solutions of $\mathbf{F}_{\mathrm{RF}}$, $\mathbf{F}_{\mathrm{BB}}[k], \forall k$ *and* $\mathbf{W}_{\mathrm{RF},u}, \forall u$ *in these two problems are identical.*

The proof of Proposition 1 is relegated to the Appendix.

When $\mathbf{F}_{\mathrm{RF}}$, $\mathbf{F}_{\mathrm{BB}}[k], \forall k$, $\mathbf{W}_{\mathrm{RF},u}, \forall u$, and $\mathbf{\Lambda}_u[k], \forall k, u$ are fixed, we can readily observe that the sub-problem with respect to $\mathbf{W}_{\mathrm{BB},u}[k]$ is a convex unconstrained problem. By setting the first-order partial derivative of the objective function in problem (11) with respect to $\mathbf{W}_{\mathrm{BB},u}[k]$ to zero, we can derive the optimal digital combiner as

$$\begin{aligned} \mathbf{W}_{\mathrm{BB},u}[k] = & \Bigg(\mathbf{W}_{\mathrm{RF},u}^H \mathbf{H}_u[k] \Bigg(\sum_{i=1}^{U} \mathbf{F}_{\mathrm{RF}} \mathbf{F}_{\mathrm{BB},i}[k] \mathbf{F}_{\mathrm{BB},i}^H[k] \\ & \times \mathbf{F}_{\mathrm{RF}}^H \Bigg) \mathbf{H}_u^H[k] \mathbf{W}_{\mathrm{RF},u} + \sigma_n^2 \mathbf{W}_{\mathrm{RF},u}^H \mathbf{W}_{\mathrm{RF},u} \Bigg)^{-1} \\ & \times \mathbf{W}_{\mathrm{RF},u}^H \mathbf{H}_u[k] \mathbf{F}_{\mathrm{RF}} \mathbf{F}_{\mathrm{BB},u}[k], \quad \forall k, u. \end{aligned} \tag{12}$$

The optimal digital combiner is also referred to as the receive Wiener filter [34], which can minimize the MSE of each user at each subcarrier independently.

Similarly, fixing $\mathbf{F}_{\mathrm{RF}}$, $\mathbf{F}_{\mathrm{BB}}[k], \forall k$, $\mathbf{W}_{\mathrm{RF},u}, \forall u$, and $\mathbf{W}_{\mathrm{BB},u}[k], \forall k, u$, the objective function of problem (11) is convex with respect to $\mathbf{\Lambda}_u[k]$. Therefore, by checking the first-order optimality condition, we obtain the optimal weight matrix $\mathbf{\Lambda}_u[k]$ as

$$\mathbf{\Lambda}_u[k] = (\mathbf{E}_u[k])^{-1}, \quad \forall k, u. \tag{13}$$

So far, we have obtained the optimal closed-form solutions of the digital combiner and weight matrix according to (12) and (13). In the following subsections, we design the analog precoder, analog combiner, and digital precoder, respectively.

B. Majorization Minimization (MM) Based Analog Precoder and Combiner Design

To deal with the non-convex constant modulus constraint, we apply the MM method to design the analog precoder and combiner in this subsection.

1) MM Method: The MM method is an effective optimization framework for tackling non-convex problems. Instead of directly dealing with the intractable objective function, the key idea of MM method is to first construct an appropriate surrogate function as the upper bound of the original objective function, then minimize the surrogate function [35].

Specifically, consider the following optimization problem

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \in \mathcal{X}, \end{aligned} \tag{14}$$

where $\mathcal{X} \subseteq \mathbb{R}^n$ is a non-empty closed set. In the majorization stage, a continuous surrogate function $g(\mathbf{x}; \mathbf{x}^{(t)})$ at the t-th iteration is constructed as the majorizer of $f(\mathbf{x})$ at the point $\mathbf{x}^{(t)}$. To guarantee that the MM method converges to a stationary point of problem (14), the surrogate function $g(\mathbf{x}; \mathbf{x}^{(t)})$ must satisfy the following three properties [35], [36], [37]

$$g(\mathbf{x}; \mathbf{x}^{(t)}) \ge f(\mathbf{x}), \tag{15a}$$

$$g(\mathbf{x}^{(t)}; \mathbf{x}^{(t)}) = f(\mathbf{x}^{(t)}), \tag{15b}$$

$$g'(\mathbf{x}^{(t)}; \mathbf{x}^{(t)}) = f'(\mathbf{x}^{(t)}). \tag{15c}$$

The above properties suggest that the surrogate function is a tight upper bound of the original objective function. In the minimization stage, $\mathbf{x}$ is updated by minimizing the surrogate function $g(\mathbf{x}; \mathbf{x}^{(t)})$, i.e.,

$$\mathbf{x}^{(t+1)} \in \arg\min_{\mathbf{x} \in \mathcal{X}} g(\mathbf{x}; \mathbf{x}^{(t)}). \tag{16}$$

According to (15) and (16), the objective function decreases monotonically at each iteration, i.e.,

$$f(\mathbf{x}^{(t+1)}) \le g(\mathbf{x}^{(t+1)}; \mathbf{x}^{(t)}) \le g(\mathbf{x}^{(t)}; \mathbf{x}^{(t)}) = f(\mathbf{x}^{(t)}), \quad \forall t. \tag{17}$$

2) MM Based Analog Precoder: The analog precoder at the BS is designed by jointly considering all users and all subcarriers. Specifically, when $\mathbf{F}_{\mathrm{BB}}[k], \forall k$, $\mathbf{W}_{\mathrm{RF},u}, \forall u$, $\mathbf{W}_{\mathrm{BB},u}[k], \forall k, u$, and $\mathbf{\Lambda}_u[k], \forall k, u$ are fixed, as well as the terms independent of $\mathbf{F}_{\mathrm{RF}}$ are omitted, problem (11) can be rewritten as a sub-problem with respect to $\mathbf{F}_{\mathrm{RF}}$ as follows

$$\begin{aligned} \min_{\mathbf{F}_{\mathrm{RF}}} \quad & \frac{1}{K} \sum_{k=1}^{K} \Big(\mathrm{Tr}\left(\mathbf{F}_{\mathrm{RF}}^H \mathbf{A}_{\mathrm{F}}[k] \mathbf{F}_{\mathrm{RF}} \mathbf{C}_{\mathrm{F}}[k] \right) \\ & - 2\Re\left\{ \mathrm{Tr}\left(\mathbf{F}_{\mathrm{RF}}^H \mathbf{B}_{\mathrm{F}}[k] \right) \right\} \Big) \\ \text{s.t.} \quad & \left| [\mathbf{F}_{\mathrm{RF}}]_{i,j} \right| = \frac{1}{\sqrt{N_t}}, \quad \forall i, j, \end{aligned} \tag{18}$$

where we define

$$\mathbf{A}_{\mathrm{F}}[k] \triangleq \mathbf{H}^H[k] \mathbf{W}_{\mathrm{RF}} \mathbf{W}_{\mathrm{BB}}[k] \mathbf{\Lambda}[k] \mathbf{W}_{\mathrm{BB}}^H[k] \mathbf{W}_{\mathrm{RF}}^H \mathbf{H}[k], \tag{19a}$$

$$\mathbf{B}_{\mathrm{F}}[k] \triangleq \mathbf{H}^H[k] \mathbf{W}_{\mathrm{RF}} \mathbf{W}_{\mathrm{BB}}[k] \mathbf{\Lambda}[k] \mathbf{F}_{\mathrm{BB}}^H[k], \tag{19b}$$

$$\mathbf{C}_{\mathrm{F}}[k] \triangleq \mathbf{F}_{\mathrm{BB}}[k] \mathbf{F}_{\mathrm{BB}}^H[k], \tag{19c}$$

with $\mathbf{H}[k] = \left[\mathbf{H}_1^T[k], \ldots, \mathbf{H}_U^T[k] \right]^T$ representing the aggregate channel matrix for all U users at the k-th subcarrier. $\mathbf{W}_{\mathrm{RF}} = \mathrm{blkdiag}\left\{ \mathbf{W}_{\mathrm{RF},1}, \ldots, \mathbf{W}_{\mathrm{RF},U} \right\}$,

$\mathbf{W}_{\mathrm{BB}}[k] = \mathrm{blkdiag}\{\mathbf{W}_{\mathrm{BB},1}[k],\ldots,\mathbf{W}_{\mathrm{BB},U}[k]\}$, and $\mathbf{\Lambda}[k] = \mathrm{blkdiag}\{\mathbf{\Lambda}_1[k],\ldots,\mathbf{\Lambda}_U[k]\}$ are block diagonal matrices. We observe that problem (18) is a quadratic programming problem with respect to $\mathbf{F}_{\mathrm{RF}}$, subject to the non-convex constant modulus constraint. By using the matrix identities $\mathrm{Tr}(\mathbf{ABCD}) = (\mathrm{vec}(\mathbf{A}^T))^T(\mathbf{D}^T \otimes \mathbf{B})\mathrm{vec}(\mathbf{C})$ and $\mathrm{Tr}(\mathbf{A}^H\mathbf{B}) = (\mathrm{vec}(\mathbf{A}))^H\mathrm{vec}(\mathbf{B})$, the objective function of problem (18) can be mathematically rewritten as the following equivalent form

$$\begin{aligned} f_{\mathrm{obj}}(\mathbf{f}_{\mathrm{RF}}) &= \frac{1}{K}\sum_{k=1}^{K}\left(\mathbf{f}_{\mathrm{RF}}^H\mathbf{M}[k]\mathbf{f}_{\mathrm{RF}} - 2\Re\left\{\mathbf{f}_{\mathrm{RF}}^H\mathbf{p}[k]\right\}\right) \\ &= \mathbf{f}_{\mathrm{RF}}^H\bar{\mathbf{M}}\mathbf{f}_{\mathrm{RF}} - 2\Re\left\{\mathbf{f}_{\mathrm{RF}}^H\bar{\mathbf{p}}\right\}, \end{aligned} \tag{20}$$

where we define that $\mathbf{f}_{\mathrm{RF}} \triangleq \mathrm{vec}(\mathbf{F}_{\mathrm{RF}})$, $\mathbf{M}[k] \triangleq (\mathbf{C}_{\mathrm{F}}[k])^T \otimes \mathbf{A}_{\mathrm{F}}[k]$, $\mathbf{p}[k] \triangleq \mathrm{vec}\left(\mathbf{B}_{\mathrm{F}}[k]\right)$, $\bar{\mathbf{M}} \triangleq \frac{1}{K}\sum_{k=1}^{K}\mathbf{M}[k]$, and $\bar{\mathbf{p}} \triangleq \frac{1}{K}\sum_{k=1}^{K}\mathbf{p}[k]$. To solve the non-convex optimization problem, we attempt to seek an appropriate majorizer of objective function (20) by exploiting the MM method. In the following, we give Lemma 1 [36] for constructing a majorizer.

Lemma 1: For any two Hermitian matrices $\mathbf{S} \in \mathbb{C}^{N\times N}$ *and* $\mathbf{T} \in \mathbb{C}^{N\times N}$ *such that* $\mathbf{T} \succeq \mathbf{S}$, *a majorizer of the quadratic function* $\mathbf{x}^H\mathbf{S}\mathbf{x}$ *at any point* $\mathbf{x}^{(t)} \in \mathbb{C}^{N\times 1}$ *is* $\mathbf{x}^H\mathbf{T}\mathbf{x} + 2\Re\left\{\mathbf{x}^H(\mathbf{S}-\mathbf{T})\mathbf{x}^{(t)}\right\} + (\mathbf{x}^{(t)})^H(\mathbf{T}-\mathbf{S})\mathbf{x}^{(t)}$.

Based on Lemma 1, the majorizer $f_{\mathrm{maj}}(\mathbf{f}_{\mathrm{RF}};\mathbf{f}_{\mathrm{RF}}^{(t)})$ of $f_{\mathrm{obj}}(\mathbf{f}_{\mathrm{RF}})$ at the point $\mathbf{f}_{\mathrm{RF}}^{(t)}$ can be represented as

$$\begin{aligned} f_{\mathrm{maj}}(\mathbf{f}_{\mathrm{RF}};\mathbf{f}_{\mathrm{RF}}^{(t)}) &= \lambda_{max}(\bar{\mathbf{M}})\mathbf{f}_{\mathrm{RF}}^H\mathbf{f}_{\mathrm{RF}} + 2\Re\left\{\mathbf{f}_{\mathrm{RF}}^H\tilde{\mathbf{p}}\right\} \\ &\quad + (\mathbf{f}_{\mathrm{RF}}^{(t)})^H(\lambda_{max}(\bar{\mathbf{M}})\mathbf{I}_{N_tN_t^{\mathrm{RF}}} - \bar{\mathbf{M}})\mathbf{f}_{\mathrm{RF}}^{(t)}, \end{aligned} \tag{21}$$

where $\lambda_{max}(\bar{\mathbf{M}})$ represents the maximum eigenvalue of $\bar{\mathbf{M}}$, $\tilde{\mathbf{p}} = (\bar{\mathbf{M}} - \lambda_{max}(\bar{\mathbf{M}})\mathbf{I}_{N_tN_t^{\mathrm{RF}}})\mathbf{f}_{\mathrm{RF}}^{(t)} - \bar{\mathbf{p}}$. It can be readily observed in (21) that the first term on the right side is constant, the last term is independent of $\mathbf{f}_{\mathrm{RF}}$, the majorizer $f_{\mathrm{maj}}(\mathbf{f}_{\mathrm{RF}};\mathbf{f}_{\mathrm{RF}}^{(t)})$ only depends on the second term. Therefore, the majorized sub-problem with respect to $\mathbf{f}_{\mathrm{RF}}$ can be simplified as

$$\begin{aligned} &\min_{\mathbf{f}_{\mathrm{RF}}} \ \Re\left\{\mathbf{f}_{\mathrm{RF}}^H\tilde{\mathbf{p}}\right\} \\ &\text{s.t.} \ \left|[\mathbf{f}_{\mathrm{RF}}]_p\right| = \frac{1}{\sqrt{N_t}}, \quad \forall p = 1,\ldots,N_tN_t^{\mathrm{RF}}, \end{aligned} \tag{22}$$

thus the semi closed-form solution of $\mathbf{f}_{\mathrm{RF}}$ can be obtained by

$$\mathbf{f}_{\mathrm{RF}} = -\frac{1}{\sqrt{N_t}}e^{j\arg(\tilde{\mathbf{p}})}. \tag{23}$$

By the inversion of matrix vectorization, the analog precoder can be expressed as

$$\mathbf{F}_{\mathrm{RF}} = \mathrm{unvec}_{N_t,N_t^{\mathrm{RF}}}(\mathbf{f}_{\mathrm{RF}}). \tag{24}$$

The steps of obtaining the analog precoder by using the MM method are summarized in **Algorithm 1**.

3) MM Based Analog Combiner: It is worth pointing out that the analog combiners of different users are independent, hence $\mathbf{W}_{\mathrm{RF},u}, \forall u$ can be implemented in parallel. Fixing $\mathbf{F}_{\mathrm{RF}}$, $\mathbf{F}_{\mathrm{BB}}[k], \forall k$, $\mathbf{W}_{\mathrm{BB},u}[k], \forall k,u$, and $\mathbf{\Lambda}_u[k], \forall k,u$ and omitting terms not relevant to $\mathbf{W}_{\mathrm{RF},u}, \forall u$, the sub-problem with respect to $\mathbf{W}_{\mathrm{RF},u}$ can be reformulated as

$$\begin{aligned} \min_{\mathbf{W}_{\mathrm{RF},u}} \ & \frac{1}{K}\sum_{k=1}^{K}\Big(\mathrm{Tr}\left(\mathbf{W}_{\mathrm{RF},u}^H\mathbf{A}_{\mathrm{W},u}[k]\mathbf{W}_{\mathrm{RF},u}\mathbf{C}_{\mathrm{W},u}[k]\right) \\ & - 2\Re\left\{\mathrm{Tr}\left(\mathbf{W}_{\mathrm{RF},u}^H\mathbf{B}_{\mathrm{W},u}[k]\right)\right\}\Big) \\ \text{s.t.} \ & \left|[\mathbf{W}_{\mathrm{RF},u}]_{m,n}\right| = \frac{1}{\sqrt{N_r}}, \quad \forall m,n, \end{aligned} \tag{25}$$

where we define

$$\mathbf{A}_{\mathrm{W},u}[k] \triangleq \mathbf{H}_u[k]\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}[k]\mathbf{F}_{\mathrm{BB}}^H[k]\mathbf{F}_{\mathrm{RF}}^H\mathbf{H}_u^H[k] + \sigma_n^2\mathbf{I}_{N_r}, \tag{26a}$$

$$\mathbf{B}_{\mathrm{W},u}[k] \triangleq \mathbf{H}_u[k]\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB},u}[k]\mathbf{\Lambda}_u[k]\mathbf{W}_{\mathrm{BB},u}^H[k], \tag{26b}$$

$$\mathbf{C}_{\mathrm{W},u}[k] \triangleq \mathbf{W}_{\mathrm{BB},u}[k]\mathbf{\Lambda}_u[k]\mathbf{W}_{\mathrm{BB},u}^H[k]. \tag{26c}$$

It can be readily observed that problem (25) is also a quadratic programming problem with respect to $\mathbf{W}_{\mathrm{RF},u}$, subject to the non-convex constant modulus constraint. According to the matrix identities $\mathrm{Tr}(\mathbf{ABCD}) = (\mathrm{vec}(\mathbf{A}^T))^T(\mathbf{D}^T \otimes \mathbf{B})\mathrm{vec}(\mathbf{C})$ and $\mathrm{Tr}(\mathbf{A}^H\mathbf{B}) = (\mathrm{vec}(\mathbf{A}))^H\mathrm{vec}(\mathbf{B})$, the objective function of problem (25) can be equivalently rewritten as

$$\begin{aligned} w_{\mathrm{obj}}(\mathbf{w}_{\mathrm{RF},u}) &= \frac{1}{K}\sum_{k=1}^{K}\Big(\mathbf{w}_{\mathrm{RF},u}^H\mathbf{N}_u[k]\mathbf{w}_{\mathrm{RF},u} \\ &\quad - 2\Re\left\{\mathbf{w}_{\mathrm{RF},u}^H\mathbf{q}_u[k]\right\}\Big) \\ &= \mathbf{w}_{\mathrm{RF},u}^H\bar{\mathbf{N}}_u\mathbf{w}_{\mathrm{RF},u} - 2\Re\left\{\mathbf{w}_{\mathrm{RF},u}^H\bar{\mathbf{q}}_u\right\}, \end{aligned} \tag{27}$$

where we define that $\mathbf{w}_{\mathrm{RF},u} \triangleq \mathrm{vec}(\mathbf{W}_{\mathrm{RF},u})$, $\mathbf{N}_u[k] \triangleq (\mathbf{C}_{\mathrm{W},u}[k])^T \otimes \mathbf{A}_{\mathrm{W},u}[k]$, $\mathbf{q}_u[k] \triangleq \mathrm{vec}\left(\mathbf{B}_{\mathrm{W},u}[k]\right)$, $\bar{\mathbf{N}}_u \triangleq \frac{1}{K}\sum_{k=1}^{K}\mathbf{N}_u[k]$, and $\bar{\mathbf{q}}_u \triangleq \frac{1}{K}\sum_{k=1}^{K}\mathbf{q}_u[k]$.

Based on Lemma 1, the majorizer $w_{\mathrm{maj}}(\mathbf{w}_{\mathrm{RF},u};\mathbf{w}_{\mathrm{RF},u}^{(t)})$ of $w_{\mathrm{obj}}(\mathbf{w}_{\mathrm{RF},u})$ at the point $\mathbf{w}_{\mathrm{RF},u}^{(t)}$ can be expressed as

$$\begin{aligned} &w_{\mathrm{maj}}(\mathbf{w}_{\mathrm{RF},u};\mathbf{w}_{\mathrm{RF},u}^{(t)}) \\ &= \lambda_{max}(\bar{\mathbf{N}}_u)\mathbf{w}_{\mathrm{RF},u}^H\mathbf{w}_{\mathrm{RF},u} + 2\Re\left\{\mathbf{w}_{\mathrm{RF},u}^H\tilde{\mathbf{q}}_u\right\} \\ &\quad + (\mathbf{w}_{\mathrm{RF},u}^{(t)})^H(\lambda_{max}(\bar{\mathbf{N}}_u)\mathbf{I}_{N_rN_r^{\mathrm{RF}}} - \bar{\mathbf{N}}_u)\mathbf{w}_{\mathrm{RF},u}^{(t)}, \end{aligned} \tag{28}$$

where $\lambda_{max}(\bar{\mathbf{N}}_u)$ represents the maximum eigenvalue of $\bar{\mathbf{N}}_u$, $\tilde{\mathbf{q}}_u = (\bar{\mathbf{N}}_u - \lambda_{max}(\bar{\mathbf{N}}_u)\mathbf{I}_{N_rN_r^{\mathrm{RF}}})\mathbf{w}_{\mathrm{RF},u}^{(t)} - \bar{\mathbf{q}}_u$. It is easily observed in (28) that, the majorizer $w_{\mathrm{maj}}(\mathbf{w}_{\mathrm{RF},u};\mathbf{w}_{\mathrm{RF},u}^{(t)})$ only relies on the second term of the right side due to the fact that the first term is constant and the last term is independent of

Algorithm 1 MM Based Analog Precoder Design

1: **Input:** $\mathbf{f}_{\mathrm{RF}}^{(0)}, \mathbf{F}_{\mathrm{BB}}[k], \mathbf{W}_{\mathrm{RF},u}, \mathbf{W}_{\mathrm{BB},u}[k], \mathbf{\Lambda}_u[k]$
2: **Initialization:** iteration index $t = 0$, convergence threshold ϵ
3: **repeat**
4: $t = t + 1$
5: Calculate $\tilde{\mathbf{p}} = (\bar{\mathbf{M}} - \lambda_{max}(\bar{\mathbf{M}})\mathbf{I}_{N_tN_t^{\mathrm{RF}}})\mathbf{f}_{\mathrm{RF}}^{(t-1)} - \bar{\mathbf{p}}$
6: Update the analog precoder vector $\mathbf{f}_{\mathrm{RF}}^{(t)}$ according to (23)
7: **until** $\left|f_{\mathrm{obj}}(\mathbf{f}_{\mathrm{RF}}^{(t)}) - f_{\mathrm{obj}}(\mathbf{f}_{\mathrm{RF}}^{(t-1)})\right| < \epsilon$
8: **Output:** $\mathbf{F}_{\mathrm{RF}}$

$\mathbf{w}_{\mathrm{RF},u}$. Therefore, the majorized sub-problem with respect to $\mathbf{w}_{\mathrm{RF},u}$ can be simplified as

$$\min_{\mathbf{w}_{\mathrm{RF},u}} \ \Re\left\{\mathbf{w}_{\mathrm{RF},u}^H \tilde{\mathbf{q}}_u\right\} \quad \text{s.t.} \ \left|[\mathbf{w}_{\mathrm{RF},u}]_q\right| = \frac{1}{\sqrt{N_r}}, \quad \forall q = 1, \ldots, N_r N_r^{\mathrm{RF}}, \tag{29}$$

hence the semi closed-form solution of $\mathbf{w}_{\mathrm{RF},u}$ can be given by

$$\mathbf{w}_{\mathrm{RF},u} = -\frac{1}{\sqrt{N_r}} e^{j \arg(\tilde{\mathbf{q}}_u)}, \quad \forall u. \tag{30}$$

After the inversion of matrix vectorization, the analog combiner can be represented as

$$\mathbf{W}_{\mathrm{RF},u} = \mathrm{unvec}_{N_r, N_r^{\mathrm{RF}}}(\mathbf{w}_{\mathrm{RF},u}), \quad \forall u. \tag{31}$$

The MM based analog combiner desgin is similar to **Algorithm 1**.

C. Element-Wise Block Coordinate Descent (EBCD) Based Analog Precoder and Combiner Design

In this subsection, we develop an EBCD algorithm for designing the analog precoder and combiner. Specifically, the sub-problem with respect to $\mathbf{F}_{\mathrm{RF}}$ can be formulated as the same form as problem (18). Inspired by the block coordinate descent method in [23] and [38], the analog precoder is designed element-by-element, i.e., only one element is updated with the other elements fixed in each step. For notational convenience, we define

$$\phi\left(\mathbf{F}_{\mathrm{RF}}\right) \triangleq \sum_{k=1}^{K} \Big(\mathrm{Tr}\left(\mathbf{F}_{\mathrm{RF}}^H \mathbf{A}_{\mathrm{F}}[k] \mathbf{F}_{\mathrm{RF}} \mathbf{C}_{\mathrm{F}}[k]\right) - 2\Re\left\{\mathrm{Tr}(\mathbf{F}_{\mathrm{RF}}^H \mathbf{B}_{\mathrm{F}}[k])\right\} \Big). \tag{32}$$

Without loss of generality, we observe that $\phi\left(\mathbf{F}_{\mathrm{RF}}\right)$ can be expressed as a quadratic function with respect to the (i,j)-th element of $\mathbf{F}_{\mathrm{RF}}$ by omitting the constant terms, which is defined as

$$\tilde{\phi}\left([\mathbf{F}_{\mathrm{RF}}]_{i,j}\right) \triangleq a\left|[\mathbf{F}_{\mathrm{RF}}]_{i,j}\right|^2 - 2\Re\left\{b^* [\mathbf{F}_{\mathrm{RF}}]_{i,j}\right\}, \tag{33}$$

where a is a real number, b is a complex number. Since each element in $\mathbf{F}_{\mathrm{RF}}$ satisfies the constant modulus constraint, the first term of $\tilde{\phi}\left([\mathbf{F}_{\mathrm{RF}}]_{i,j}\right)$ is a constant. Therefore, the sub-problem with respect to $[\mathbf{F}_{\mathrm{RF}}]_{i,j}$ can be represented as

$$\max_{[\mathbf{F}_{\mathrm{RF}}]_{i,j}} \ \Re\left\{b^* [\mathbf{F}_{\mathrm{RF}}]_{i,j}\right\} \quad \text{s.t.} \ \left|[\mathbf{F}_{\mathrm{RF}}]_{i,j}\right| = \frac{1}{\sqrt{N_t}}. \tag{34}$$

It is easy to find that the optimal solution of problem (34) is given by $[\mathbf{F}_{\mathrm{RF}}]_{i,j} = \frac{1}{\sqrt{N_t}} e^{j \arg(b)}$. To update $[\mathbf{F}_{\mathrm{RF}}]_{i,j}$, we only need to know the value of b. Next, we give an approach for obtaining the complex number b. According to [23], the following equation holds

$$\left.\frac{\partial \tilde{\phi}\left([\mathbf{F}_{\mathrm{RF}}]_{i,j}\right)}{\partial [\mathbf{F}_{\mathrm{RF}}]_{i,j}^*}\right|_{[\mathbf{F}_{\mathrm{RF}}]_{i,j} = [\tilde{\mathbf{F}}_{\mathrm{RF}}]_{i,j}} = \frac{1}{2}\left(a\left[\tilde{\mathbf{F}}_{\mathrm{RF}}\right]_{i,j} - b\right). \tag{35}$$

Algorithm 2 EBCD Based Analog Precoder Desgin

Input: $\mathbf{F}_{\mathrm{RF}}^{(0)}, \mathbf{A}_{\mathrm{F}}[k], \mathbf{B}_{\mathrm{F}}[k], \mathbf{C}_{\mathrm{F}}[k]$
Initialization: $\mathbf{Z}^{(0)}[k] = \mathbf{A}_{\mathrm{F}}[k]\mathbf{F}_{\mathrm{RF}}^{(0)}\mathbf{C}_{\mathrm{F}}[k], \forall k$, iteration index $t = 0$, convergence threshold ϵ
repeat
  $t = t + 1$
  for $i = 1$ **to** N_t **do**
    for $j = 1$ **to** N_t^{RF} **do**
      $b = \sum_{k=1}^{K} \Big([\mathbf{A}_{\mathrm{F}}[k]]_{i,i} \left[\mathbf{F}_{\mathrm{RF}}^{(t-1)}\right]_{i,j} [\mathbf{C}_{\mathrm{F}}[k]]_{j,j} - \left[\mathbf{Z}^{(t-1)}[k]\right]_{i,j} + [\mathbf{B}_{\mathrm{F}}[k]]_{i,j} \Big)$
      $x = \frac{1}{\sqrt{N_t}} e^{j \arg(b)}$
      for $k = 1$ **to** K **do**
        $\mathbf{Z}^{(t)}[k] = \mathbf{Z}^{(t-1)}[k] + \left(x - \left[\mathbf{F}_{\mathrm{RF}}^{(t-1)}\right]_{i,j}\right) [\mathbf{A}_{\mathrm{F}}[k]]_{:,i} [\mathbf{C}_{\mathrm{F}}[k]]_{j,:}$
      end for
      $\left[\mathbf{F}_{\mathrm{RF}}^{(t)}\right]_{i,j} = x$
    end for
  end for
until $\left|\phi\left(\mathbf{F}_{\mathrm{RF}}^{(t)}\right) - \phi\left(\mathbf{F}_{\mathrm{RF}}^{(t-1)}\right)\right| < \epsilon$
Output: $\mathbf{F}_{\mathrm{RF}} = \mathbf{F}_{\mathrm{RF}}^{(t)}$

Besides, based on [23], we have

$$\left.\frac{\partial \phi\left(\mathbf{F}_{\mathrm{RF}}\right)}{\partial \mathbf{F}_{\mathrm{RF}}^*}\right|_{\mathbf{F}_{\mathrm{RF}} = \tilde{\mathbf{F}}_{\mathrm{RF}}} = \frac{1}{2}\sum_{k=1}^{K}\left(\mathbf{A}_{\mathrm{F}}[k]\tilde{\mathbf{F}}_{\mathrm{RF}}\mathbf{C}_{\mathrm{F}}[k] - \mathbf{B}_{\mathrm{F}}[k]\right). \tag{36}$$

Combining (35) and (36), we can further obtain

$$\sum_{k=1}^{K}\left[\mathbf{A}_{\mathrm{F}}[k]\tilde{\mathbf{F}}_{\mathrm{RF}}\mathbf{C}_{\mathrm{F}}[k] - \mathbf{B}_{\mathrm{F}}[k]\right]_{i,j} = a\left[\tilde{\mathbf{F}}_{\mathrm{RF}}\right]_{i,j} - b. \tag{37}$$

By expanding $\sum_{k=1}^{K}\left[\mathbf{A}_{\mathrm{F}}[k]\tilde{\mathbf{F}}_{\mathrm{RF}}\mathbf{C}_{\mathrm{F}}[k]\right]_{i,j}$ and checking the coefficient of $a\left[\tilde{\mathbf{F}}_{\mathrm{RF}}\right]_{i,j}$, we have

$$a\left[\tilde{\mathbf{F}}_{\mathrm{RF}}\right]_{i,j} = \sum_{k=1}^{K}[\mathbf{A}_{\mathrm{F}}[k]]_{i,i}\left[\tilde{\mathbf{F}}_{\mathrm{RF}}\right]_{i,j}[\mathbf{C}_{\mathrm{F}}[k]]_{j,j}. \tag{38}$$

Thus, b can be acquired according to the following equation

$$b = \sum_{k=1}^{K}\Big([\mathbf{A}_{\mathrm{F}}[k]]_{i,i}\left[\tilde{\mathbf{F}}_{\mathrm{RF}}\right]_{i,j}[\mathbf{C}_{\mathrm{F}}[k]]_{j,j} + [\mathbf{B}_{\mathrm{F}}[k]]_{i,j} - \left[\mathbf{A}_{\mathrm{F}}[k]\tilde{\mathbf{F}}_{\mathrm{RF}}\mathbf{C}_{\mathrm{F}}[k]\right]_{i,j}\Big). \tag{39}$$

Based on the above analysis, each element in $\mathbf{F}_{\mathrm{RF}}$ can be alternately updated. The proposed EBCD algorithm for the analog precoder design is summarized in **Algorithm 2**.

Furthermore, we can observe that problem (25) with respect to $\mathbf{W}_{\mathrm{RF},u}$ has the same form as problem (18) with respect to $\mathbf{F}_{\mathrm{RF}}$. Therefore, similar to the analog precoder, the analog combiner for each user can also be obtained using **Algorithm 2** by replacing the input $\mathbf{F}_{\mathrm{RF}}^{(0)}, \mathbf{A}_{\mathrm{F}}[k], \mathbf{B}_{\mathrm{F}}[k], \mathbf{C}_{\mathrm{F}}[k]$ with $\mathbf{W}_{\mathrm{RF},u}^{(0)}, \mathbf{A}_{\mathrm{W},u}[k], \mathbf{B}_{\mathrm{W},u}[k], \mathbf{C}_{\mathrm{W},u}[k]$, respectively.

D. Lagrangian Multiplier Method Based Digital Precoder

In this subsection, we design the digital precoder based on the Lagrangian multiplier method. To deal with the coupled transmit power constraint between the analog precoder $\mathbf{F}_{\mathrm{RF}}$ and digital precoder $\mathbf{F}_{\mathrm{BB}}[k], \forall k$, we introduce the effective analog precoder $\tilde{\mathbf{F}}_{\mathrm{RF}} = \mathbf{F}_{\mathrm{RF}} \left(\mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}}\right)^{-\frac{1}{2}}$ and effective digital precoder $\tilde{\mathbf{F}}_{\mathrm{BB}}[k] = \left(\mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}}\right)^{\frac{1}{2}} \mathbf{F}_{\mathrm{BB}}[k]$ such that $\tilde{\mathbf{F}}_{\mathrm{BB}}[k]$ satisfies the power constraint independently, i.e., $\left\|\tilde{\mathbf{F}}_{\mathrm{BB}}[k]\right\|_F^2 = \mathrm{Tr}\left(\mathbf{F}_{\mathrm{BB}}^H[k]\mathbf{F}_{\mathrm{RF}}^H\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}[k]\right) = \left\|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}[k]\right\|_F^2 \leq P_k$. Therefore, we consider the closed-form solution of $\tilde{\mathbf{F}}_{\mathrm{BB}}[k]$ at each subcarrier in the following.

Specifically, when $\mathbf{F}_{\mathrm{RF}}$, $\mathbf{W}_{\mathrm{RF},u}, \forall u$, $\mathbf{W}_{\mathrm{BB},u}[k], \forall k, u$, and $\mathbf{\Lambda}_u[k], \forall k, u$ are fixed, the sub-problem with respect to the effective digital precoder $\tilde{\mathbf{F}}_{\mathrm{BB}}[k]$ can be expressed as

$$\begin{aligned}\min_{\tilde{\mathbf{F}}_{\mathrm{BB}}[k]} \ & \mathrm{Tr}\left(\tilde{\mathbf{F}}_{\mathrm{BB}}^H[k]\tilde{\mathbf{A}}_{\mathrm{F}}[k]\tilde{\mathbf{F}}_{\mathrm{BB}}[k]\right) \\ & - 2\Re\left\{\mathrm{Tr}\left(\tilde{\mathbf{F}}_{\mathrm{BB}}^H[k]\tilde{\mathbf{B}}_{\mathrm{F}}[k]\right)\right\} \\ \text{s.t.} \ & \left\|\tilde{\mathbf{F}}_{\mathrm{BB}}[k]\right\|_F^2 \leq P_k, \end{aligned} \tag{40}$$

where we define

$$\begin{aligned}\tilde{\mathbf{A}}_{\mathrm{F}}[k] \triangleq \ & \tilde{\mathbf{F}}_{\mathrm{RF}}^H\mathbf{H}^H[k]\mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}[k]\mathbf{\Lambda}[k]\mathbf{W}_{\mathrm{BB}}^H[k] \\ & \times \mathbf{W}_{\mathrm{RF}}^H\mathbf{H}[k]\tilde{\mathbf{F}}_{\mathrm{RF}}, \end{aligned} \tag{41a}$$

$$\tilde{\mathbf{B}}_{\mathrm{F}}[k] \triangleq \tilde{\mathbf{F}}_{\mathrm{RF}}^H\mathbf{H}^H[k]\mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}[k]\mathbf{\Lambda}[k]. \tag{41b}$$

It is evidently observed that problem (40) is a convex quadratically constrained quadratic programming (QCQP) problem, which can be solved using the Lagrangian multiplier method [31]. By introducing a Lagrangian multiplier $\mu_k \geq 0$ to the power constraint at each subcarrier, problem (40) can be transformed into an unconstrained quadratic programming problem as

$$\begin{aligned}\min_{\tilde{\mathbf{F}}_{\mathrm{BB}}[k]} \ & \mathrm{Tr}\left(\tilde{\mathbf{F}}_{\mathrm{BB}}^H[k]\tilde{\mathbf{A}}_{\mathrm{F}}[k]\tilde{\mathbf{F}}_{\mathrm{BB}}[k]\right) - 2\Re\left\{\mathrm{Tr}\left(\tilde{\mathbf{F}}_{\mathrm{BB}}^H[k]\tilde{\mathbf{B}}_{\mathrm{F}}[k]\right)\right\} \\ & + \mu_k\left(\left\|\tilde{\mathbf{F}}_{\mathrm{BB}}[k]\right\|_F^2 - P_k\right). \end{aligned} \tag{42}$$

By setting the partial derivative of the objective function in problem (42) with respect to $\tilde{\mathbf{F}}_{\mathrm{BB}}[k]$ to zero, the optimal effective digital precoder can be obtained by

$$\tilde{\mathbf{F}}_{\mathrm{BB}}[k] = \left(\tilde{\mathbf{A}}_{\mathrm{F}}[k] + \mu_k\mathbf{I}_{N_t^{\mathrm{RF}}}\right)^{-1}\tilde{\mathbf{B}}_{\mathrm{F}}[k], \tag{43}$$

where the optimal value of μ_k is chosen such that the complementary slackness condition is satisfied, i.e., $\mu_k\left(\left\|\tilde{\mathbf{F}}_{\mathrm{BB}}[k]\right\|_F^2 - P_k\right) = 0$. To determine the optimal value of μ_k, the transmit power of $\tilde{\mathbf{F}}_{\mathrm{BB}}[k]$ can be rewritten as

$$\left\|\tilde{\mathbf{F}}_{\mathrm{BB}}[k]\right\|_F^2 = \mathrm{Tr}\left(\tilde{\mathbf{B}}_{\mathrm{F}}^H[k]\left(\tilde{\mathbf{A}}_{\mathrm{F}}[k] + \mu_k\mathbf{I}_{N_t^{\mathrm{RF}}}\right)^{-2}\tilde{\mathbf{B}}_{\mathrm{F}}[k]\right) \tag{44a}$$

$$\overset{(a)}{=} \mathrm{Tr}\left(\left(\tilde{\mathbf{A}}_{\mathrm{F}}[k] + \mu_k\mathbf{I}_{N_t^{\mathrm{RF}}}\right)^{-2}\mathbf{\Psi}[k]\right) \tag{44b}$$

$$\overset{(b)}{=} \mathrm{Tr}\left(\left(\mathbf{U}[k]\mathbf{\Sigma}[k]\mathbf{U}^H[k] + \mu_k\mathbf{I}_{N_t^{\mathrm{RF}}}\right)^{-2}\mathbf{\Psi}[k]\right) \tag{44c}$$

$$\overset{(c)}{=} \mathrm{Tr}\left(\left(\mathbf{\Sigma}[k] + \mu_k\mathbf{I}_{N_t^{\mathrm{RF}}}\right)^{-2}\bar{\mathbf{\Psi}}[k]\right) \tag{44d}$$

$$\overset{(d)}{=} \sum_{n=1}^{N_t^{\mathrm{RF}}} \frac{\left[\bar{\mathbf{\Psi}}[k]\right]_{n,n}}{\left(\left[\mathbf{\Sigma}[k]\right]_{n,n} + \mu_k\right)^2}, \tag{44e}$$

where (a) holds by defining $\mathbf{\Psi}[k] \triangleq \tilde{\mathbf{B}}_{\mathrm{F}}[k]\tilde{\mathbf{B}}_{\mathrm{F}}^H[k]$, (b) is obtained by performing the eigenvalue decomposition $\tilde{\mathbf{A}}_{\mathrm{F}}[k] = \mathbf{U}[k]\mathbf{\Sigma}[k]\mathbf{U}^H[k]$, (c) holds since $\mathbf{U}[k]$ is a unitary matrix and $\bar{\mathbf{\Psi}}[k] = \mathbf{U}^H[k]\mathbf{\Psi}[k]\mathbf{U}[k]$, (d) holds by defining $\left[\bar{\mathbf{\Psi}}[k]\right]_{n,n}$ and $\left[\mathbf{\Sigma}[k]\right]_{n,n}$ as the (n,n)-th element on the diagonal of $\bar{\mathbf{\Psi}}[k]$ and $\mathbf{\Sigma}[k]$, respectively. It is worth noting in (44e) that $\left\|\tilde{\mathbf{F}}_{\mathrm{BB}}[k]\right\|_F^2$ is monotonically decreasing with respect to μ_k. When $\mu_k = 0$, we obtain the upper bound $\sum_{n=1}^{N_t^{\mathrm{RF}}} \frac{\left[\bar{\mathbf{\Psi}}[k]\right]_{n,n}}{\left(\left[\mathbf{\Sigma}[k]\right]_{n,n}\right)^2}$. To determine the optimal value of μ_k, we discuss the following two cases:

Case 1: If the upper bound $\sum_{n=1}^{N_t^{\mathrm{RF}}} \frac{\left[\bar{\mathbf{\Psi}}[k]\right]_{n,n}}{\left(\left[\mathbf{\Sigma}[k]\right]_{n,n}\right)^2} \leq P_k$, the transmit power constraint is naturally satisfied for all $\mu_k \geq 0$. Thus, we set the optimal μ_k to 0.

Case 2: If the upper bound $\sum_{n=1}^{N_t^{\mathrm{RF}}} \frac{\left[\bar{\mathbf{\Psi}}[k]\right]_{n,n}}{\left(\left[\mathbf{\Sigma}[k]\right]_{n,n}\right)^2} > P_k$, the optimal μ_k can be easily obtained by utilizing the bisection search method [31]. Specifically, the search range of μ_k is set to $\mu_k \in [\mu_{\mathrm{low}}, \mu_{\mathrm{up}}]$, where μ_{low} is initialized as 0 and μ_{up} is initialized as $\sqrt{\frac{\sum_{n=1}^{N_t^{\mathrm{RF}}}\left[\bar{\mathbf{\Psi}}[k]\right]_{n,n}}{P_k}}$ [39]. This is because $\sum_{n=1}^{N_t^{\mathrm{RF}}} \frac{\left[\bar{\mathbf{\Psi}}[k]\right]_{n,n}}{\left(\left[\mathbf{\Sigma}[k]\right]_{n,n}+\mu_k\right)^2} < \sum_{n=1}^{N_t^{\mathrm{RF}}} \frac{\left[\bar{\mathbf{\Psi}}[k]\right]_{n,n}}{\mu_k^2} \leq P_k$. The detailed procedure of obtaining the optimal Lagrangian multiplier by the bisection search is summarized in **Algorithm 3**.

Algorithm 3 Bisection Search for Lagrangian Multiplier

1: **Initialization:** search range $[\mu_{\mathrm{low}}, \mu_{\mathrm{up}}]$, iteration index $t = 0$, convergence threshold ϵ
2: **repeat**
3: $t = t + 1$;
4: Calculate $\mu_k = \frac{\mu_{\mathrm{low}} + \mu_{\mathrm{up}}}{2}$;
5: **if** $\left\|\tilde{\mathbf{F}}_{\mathrm{BB}}[k]\right\|_F^2 > P_k$ **then**
6: Set $\mu_{\mathrm{low}} = \mu_k$;
7: **else**
8: Set $\mu_{\mathrm{up}} = \mu_k$;
9: **end if**
10: **until** $|\mu_{\mathrm{up}} - \mu_{\mathrm{low}}| < \epsilon$
11: **Output:** μ_k

By substituting the optimal μ_k into (43), the optimal digital precoder can be expressed as

$$\begin{aligned}\mathbf{F}_{\mathrm{BB}}[k] &= \left(\mathbf{F}_{\mathrm{RF}}^H\mathbf{F}_{\mathrm{RF}}\right)^{-\frac{1}{2}}\tilde{\mathbf{F}}_{\mathrm{BB}}[k] \\ &= \left(\mathbf{F}_{\mathrm{RF}}^H\mathbf{F}_{\mathrm{RF}}\right)^{-\frac{1}{2}}\left(\tilde{\mathbf{A}}_{\mathrm{F}}[k] + \mu_k\mathbf{I}_{N_t^{\mathrm{RF}}}\right)^{-1}\tilde{\mathbf{B}}_{\mathrm{F}}[k]. \end{aligned} \tag{45}$$

Combining (12), (13), **Algorithm 1**, **Algorithm 3**, and (45), we can obtain the alternating optimization algorithm based on MM (abbreviated as "AO-MM") for hybrid transceiver design. Similarly, we acquire the alternating optimization algorithm based on EBCD (abbreviated as "AO-EBCD") by

Algorithm 4 Alternating Optimization Based on MM (or EBCD)

Input: $\{\mathbf{H}_u[k], \forall k, u\}, \{P_k, \forall k\}, \sigma_n^2$
Initialization: $\mathbf{F}_{\mathrm{RF}}^{(0)}, \left\{\mathbf{F}_{\mathrm{BB}}^{(0)}[k], \forall k\right\}, \left\{\mathbf{W}_{\mathrm{RF},u}^{(0)}, \forall u\right\}$, $g_{\mathrm{WMMSE}}^{(0)} = 0$, iteration index $n = 0$, convergence threshold ϵ
repeat
  $n = n + 1$
  Update the digital combiner $\left\{\mathbf{W}_{\mathrm{BB},u}^{(n)}[k], \forall k, u\right\}$ according to (12)
  Update the weight matrix $\left\{\mathbf{\Lambda}_u^{(n)}[k], \forall k, u\right\}$ according to (13)
  Update the analog combiner $\left\{\mathbf{W}_{\mathrm{RF},u}^{(n)}, \forall u\right\}$ using **Algorithm 1** or **Algorithm 2**
  Update the analog precoder $\mathbf{F}_{\mathrm{RF}}^{(n)}$ using **Algorithm 1** or **Algorithm 2**
  Update the digital precoder $\left\{\mathbf{F}_{\mathrm{BB}}^{(n)}[k], \forall k\right\}$ according to (45) and **Algorithm 3**
until $\left|g_{\mathrm{WMMSE}}^{(n)} - g_{\mathrm{WMMSE}}^{(n-1)}\right| < \epsilon$
Output: $\mathbf{F}_{\mathrm{RF}}$, $\{\mathbf{F}_{\mathrm{BB}}[k], \forall k\}$, $\{\mathbf{W}_{\mathrm{RF},u}, \forall u\}$, $\{\mathbf{W}_{\mathrm{BB},u}[k], \forall k, u\}$

combining (12), (13), **Algorithm 2**, **Algorithm 3**, and (45). For convenience, the detailed procedure of the AO-MM and AO-EBCD is summarized in **Algorithm 4**, where $g_{\mathrm{WMMSE}}^{(n)}$ represents the objective function value of the WMMSE problem (11) in the n-th iteration.

IV. DFT Codebook Based Low-Complexity Scheme

In this section, we propose a DFT codebook based low-complexity two-stage hybrid transceiver design. At first, by adopting the DFT codebook, the analog precoder at the BS and the analog combiners at all users are jointly designed to maximize the beamforming gain and combat the inter-beam interference in the analog domain. Then, the BD method is applied to further eliminate the multiuser interference in the digital domain.

In particular, $\mathbf{D}_t \in \mathbb{C}^{N_t \times N_t}$ and $\mathbf{D}_{r,u} \in \mathbb{C}^{N_r \times N_r}$ represent the predefined DFT codebooks for the BS and the u-th user, respectively. It is worth mentioning that the DFT codebook has two main advantages. On the one hand, all elements of DFT codebook naturally satisfy constant modulus constraint, hence it can serve as a candidate for the analog precoder and combiners. On the other hand, due to the unitary property of DFT codebook, different beam steering vectors are orthogonal to each other, thus the inter-beam interference can be effectively mitigated.

Specifically, for the u-th user, we aim to jointly select N_r^{RF} column vectors from $\mathbf{D}_t$ to construct $\mathbf{F}_{\mathrm{RF}}$ and N_r^{RF} column vectors from $\mathbf{D}_{r,u}$ to construct $\mathbf{W}_{\mathrm{RF},u}$ to maximize the beamforming gain. By exploiting the common sparsity of channels among different subcarriers in the angle domain [18], the beamforming gain maximization problem can be

Algorithm 5 DFT Codebook Based Joint Analog Precoder and Combiner Design

Input: $\{\mathbf{H}_u[k], \forall k, u\}, \mathbf{D}_t, \{\mathbf{D}_{r,u}, \forall u\}$
Initialize: $\mathbf{F}_{\mathrm{RF}} = \varnothing, \mathbf{W}_{\mathrm{RF},u} = \varnothing, \forall u,\ \mathcal{U} = \{1, \ldots, U\}$
for $u = 1 : U$ **do**
  $\mathbf{\Phi}(i, j, u) = \sum_{k=1}^{K} \left|(\mathbf{d}_{r,u}^{(i)})^H \mathbf{H}_u[k] \mathbf{d}_t^{(j)}\right|^2, \forall i, j$
end for
for $n = 1 : N_t^{\mathrm{RF}}$ **do**
  $\{u^\star, i^\star, j^\star\} = \underset{u \in \mathcal{U}, i \in 1, \ldots, N_r, j \in 1, \ldots, N_t}{\arg\max} \mathbf{\Phi}(i, j, u)$
  $\mathbf{W}_{\mathrm{RF},u^\star} = \left[\mathbf{W}_{\mathrm{RF},u^\star} \mid \mathbf{d}_{r,u^\star}^{(i^\star)}\right]$, $\mathbf{F}_{\mathrm{RF}} = \left[\mathbf{F}_{\mathrm{RF}} \mid \mathbf{d}_t^{(j^\star)}\right]$
  $\mathbf{\Phi}(i^\star, :, u^\star) = 0$, $\mathbf{\Phi}(:, j^\star, :) = 0$
  for $u = 1 : U$ **do**
    if $\mathrm{rank}(\mathbf{W}_{\mathrm{RF},u}) = N_r^{\mathrm{RF}}$ **then**
      $\mathbf{\Phi}(:, :, u^\star) = \mathbf{0}$
      $\mathcal{U} = \mathrm{setdiff}(\mathcal{U}, u^\star)$
    end if
  end for
end for
Output: $\mathbf{F}_{\mathrm{RF}}$, $\{\mathbf{W}_{\mathrm{RF},u}, \forall u\}$

formulated as

$$\begin{aligned} \max_{\mathbf{d}_{r,u}^{(i)}, \mathbf{d}_t^{(j)}} \ & \sum_{k=1}^{K} \left|(\mathbf{d}_{r,u}^{(i)})^H \mathbf{H}_u[k] \mathbf{d}_t^{(j)}\right|^2 \\ \text{s.t. } & \mathbf{d}_{r,u}^{(i)} \in \mathbf{D}_{r,u}, \quad \forall i, \\ & \mathbf{d}_t^{(j)} \in \mathbf{D}_t, \quad \forall j, \end{aligned} \tag{46}$$

where $\mathbf{d}_{r,u}^{(i)}$ represents the i-th column vector of $\mathbf{D}_{r,u}$, $\mathbf{d}_t^{(j)}$ represents the j-th column vector of $\mathbf{D}_t$. For simplicity, we define that $\mathbf{\Phi}$ is a three-dimensional array with $\mathbf{\Phi}(i, j, u) = \sum_{k=1}^{K} \left|(\mathbf{d}_{r,u}^{(i)})^H \mathbf{H}_u[k] \mathbf{d}_t^{(j)}\right|^2$. Therefore, the optimal user index and DFT codeword indices at the u-th user and the BS with the maximum beamforming gain can be searched from $\mathbf{\Phi}$ as follows

$$\{u^\star, i^\star, j^\star\} = \underset{u \in \mathcal{U}, i \in 1, \ldots, N_r, j \in 1, \ldots, N_t}{\arg\max} \mathbf{\Phi}(i, j, u). \tag{47}$$

Thus, the codewords $\mathbf{d}_{r,u^\star}^{(i^\star)}$ and $\mathbf{d}_t^{(j^\star)}$ corresponding to the optimal user index $u^\star$ are respectively assigned to $\mathbf{W}_{\mathrm{RF},u^\star}$ and $\mathbf{F}_{\mathrm{RF}}$, and then the elements in the $i^\star$-th row of $\mathbf{D}_{r,u^\star}$ and the $j^\star$-th column of $\mathbf{D}_t$ are set to zero. If the design of $\mathbf{W}_{\mathrm{RF},u^\star}$ is completed, all elements of $\mathbf{\Phi}(:, :, u^\star)$ are set to zero, i.e., $\mathbf{\Phi}(:, :, u^\star) = \mathbf{0}$, and $u^\star$ is removed from $\mathcal{U}$. The above process is performed until the design of $\mathbf{F}_{\mathrm{RF}}$ is finished. For convenience, The detailed steps of the DFT codebook based joint analog precoder and combiner design are summarized in **Algorithm 5**.

Furthermore, the BD technique [16], [40] in narrowband scenarios is extended to wideband scenarios to eliminate the multiuser interference in the digital domain.

V. Convergence and Complexity Analysis

A. Convergence Analysis

In this subsection, we investigate the convergence of the proposed AO-MM and AO-EBCD algorithms. It is observed

TABLE II
COMPUTATIONAL COMPLEXITY OF DIFFERENT HYBRID TRANSCEIVER DESIGNS FOR MMWAVE MU-MIMO-OFDM SYSTEMS

Hybrid transceiver designs	Computational complexity
Proposed AO-MM	$\mathcal{O}\left(I_{\text{out}}I_{\text{in}}KN_t^2(N_t^{\text{RF}})^2\right)$
Proposed AO-EBCD	$\mathcal{O}\left(I_{\text{out}}I_{\text{in}}KN_t^2(N_t^{\text{RF}})^2\right)$
Proposed DFT-BD	$\mathcal{O}\left(UKN_rN_t^2\right)$
PCA-BD [20]	$\mathcal{O}\left(UKN_rN_t^2+UKN_sN_t^2\right)$
AM-PE [21]	$\mathcal{O}\left(UKN_rN_t^2+I_{\text{ite}}UKN_sN_tN_t^{\text{RF}}\right)$
TD-RCD [22]	$\mathcal{O}\left(UN_sN_t[K(I_{\text{ite}}N_r+UN_r+U^2N_s^2)+I_{\text{ite}}N_t]\right)$

in **Algorithm 4** that each variable is updated with the other variables fixed in each iteration. Specifically, the optimal closed-form solutions of the digital precoder $\mathbf{F}_{\text{BB}}[k]$ and the digital combiner $\mathbf{W}_{\text{BB},u}[k]$ are derived by the Lagrangian multiplier method and the first-order optimality condition, respectively. For the analog precoder $\mathbf{F}_{\text{RF}}$ and the anglog combiner $\mathbf{W}_{\text{RF},u}$, the proposed MM method and EBCD method guarantee that the objective function of WMMSE problem (11) decreases monotonically and converges to a stationary point [23], [35]. On the contrary, the spectral efficiency in problem (9) is monotonically increasing. Moreover, the spectral efficiency is upper bounded due to the transmit power constraint. As a result, **Algorithm 4** is guaranteed to converge to a locally optimal solution of problem (9).

B. Complexity Analysis

In this subsection, we analyze the computational complexity of our proposed AO-MM, AO-EBCD, and DFT-BD algorithms, respectively. **Algorithm 4** is a double-loop structure composed of the alternating optimization among variables in the outer loop and the MM (or EBCD) based analog precoder and combiner in the inner loop. Thereofore, the complexity of **Algorithm 4** is dominant by the MM (or EBCD) based analog precoder and combiner, as well as the bisection search based digital precoder, whereas the complexity of the digital combiner and weight matrix is negligible. For convenience, let I_{out} and I_{in} represent the numbers of outer and inner iterations required for convergence, respectively.

For the AO-MM algorithm, we observe that the computational cost of the analog precoder and combiner depends on **Algorithm 1**. Specifically, the complexity of the analog precoder mainly depends on the complexity of (20), which is given by $\mathcal{O}(KN_t^2(N_t^{\text{RF}})^2)$. The analog combiner mostly relies on (27) with complexity of $\mathcal{O}(KN_r^2(N_r^{\text{RF}})^2)$. Therefore, the complexities of the MM based analog precoder and combiner are $\mathcal{O}(I_{\text{in}}KN_t^2(N_t^{\text{RF}})^2)$ and $\mathcal{O}(UI_{\text{in}}KN_r^2(N_r^{\text{RF}})^2)$, respectively. The complexity of computing the digital precoder in (45) by using bisection search is given by $\mathcal{O}\left(I_{\text{bs}}K(N_t^{\text{RF}})^3\right)$, where $I_{\text{bs}}=\log(\frac{\mu_{\text{up}}-\mu_{\text{low}}}{\epsilon})$ is the number of iterations of bisection search. Therefore, the total computational complexity of the proposed AO-MM algorithm is given by

$$\mathcal{O}\Big(I_{\text{out}}I_{\text{in}}KN_t^2(N_t^{\text{RF}})^2+I_{\text{out}}I_{\text{in}}UKN_r^2(N_r^{\text{RF}})^2+I_{\text{out}}I_{\text{bs}}K(N_t^{\text{RF}})^3\Big). \tag{48}$$

For the AO-EBCD algorithm, it is easily observed that the computational complexity mainly comes from **Algorithm 2**. To be specific, the complexity of **Algorithm 2** is dominant by its steps 9 to 11, i.e., $\mathcal{O}(KN_tN_t^{\text{RF}})$ for the analog precoder and $\mathcal{O}(KN_rN_r^{\text{RF}})$ for the analog combiner, thereby the complexities of the EBCD based analog precoder and combiner are given by $\mathcal{O}(I_{\text{in}}KN_t^2(N_t^{\text{RF}})^2)$ and $\mathcal{O}(UI_{\text{in}}KN_r^2(N_r^{\text{RF}})^2)$, respectively. The complexity of the digtial precoder in the AO-EBCD algorithm is the same as that in the AO-MM algorithm. Thus, the total computational complexity of the proposed AO-EBCD algorithm is given by

$$\mathcal{O}\Big(I_{\text{out}}I_{\text{in}}KN_t^2(N_t^{\text{RF}})^2+I_{\text{out}}I_{\text{in}}UKN_r^2(N_r^{\text{RF}})^2+I_{\text{out}}I_{\text{bs}}K(N_t^{\text{RF}})^3\Big). \tag{49}$$

For the DFT-BD scheme, the computational cost mainly depends on the complexity of steps 3 to 5 in **Algorithm 5**, which is given by $\mathcal{O}(UKN_rN_t^2+UKN_r^2N_t)$.

Due to the fact that $N_t \gg N_r$ and $N_t \gg N_t^{\text{RF}}$, the computational complexity of the AO-MM, AO-EBCD, and DFT-BD is dominant by $\mathcal{O}(I_{\text{out}}I_{\text{in}}KN_t^2(N_t^{\text{RF}})^2)$, $\mathcal{O}(I_{\text{out}}I_{\text{in}}KN_t^2(N_t^{\text{RF}})^2)$, and $\mathcal{O}(UKN_rN_t^2)$, respectively. The computational complexity of our proposed hybrid transceiver designs and the existing schemes [20], [21], [22] is shown in Table II, where I_{ite} is the number of iterations needed for designing the analog precoder and combiner in [21] and [22].

VI. SIMULATION RESULTS

This section presents simulation results to evaluate the performance of our proposed hybrid transceiver designs. To demonstrate the effectiveness of the proposed schemes, the fully digital WMMSE solution [31] and the state-of-the-art benchmarks including the PCA-BD [20], the AM-PE [21], and the TD-RCD [22] are considered for comparison. Specifically, we consider a wideband mmWave massive MU-MIMO-OFDM system with hybrid architecture at the BS and all users. Throughout this section, the simulation parameters are set as follows unless otherwise specified. The numbers of transmit antennas and RF chains at the BS are $N_t=64$ and $N_t^{\text{RF}}=8$, respectively, the number of users is $U=4$, the numbers of receive antennas and RF chains at each user are $N_r=16$ and $N_r^{\text{RF}}=2$, respectively, the number of subcarriers is $K=128$, and the number of data streams

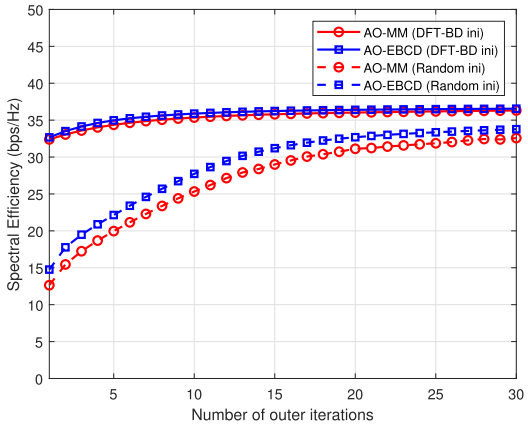

Fig. 2. Convergence of the proposed AO-MM and AO-EBCD algorithms using different initialization methods when $N_t = 64$, $N_r = 16$, $N_t^{\mathrm{RF}} = 8$, $N_r^{\mathrm{RF}} = 2$, $U = 4$, $N_s = 2$, $K = 128$, and $\mathrm{SNR} = 0$ dB.

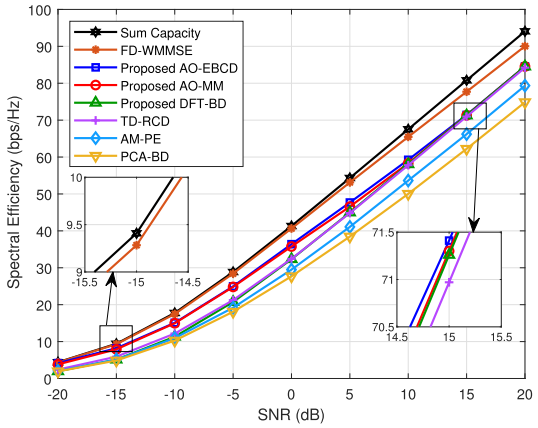

Fig. 3. Spectral efficiency versus SNR for different schemes when $N_t = 64$, $N_r = 16$, $N_t^{\mathrm{RF}} = 8$, $N_r^{\mathrm{RF}} = 2$, $U = 4$, $N_s = 2$, and $K = 128$.

transmitted from the BS to each user at each subcarrier is $N_s = 2$. Similar to [32], the bandwidth B is 200 MHz, the sampling interval is $T_s = 1/B$, the number of delay taps of the channel is $D = 16$, $\tau_{u,l}$ is uniformly distributed in $[0, DT_s]$. $p(\tau)$ is a raised-cosine filter with roll-off factor of 1 [41]. For wideband mmWave channels, it is assumed that the number of propagation paths between the BS and each user is $L_u = 10$, and the AoD $\phi_{u,l}$ and the AoA $\theta_{u,l}$ of each path are uniformly distributed in $[0, 2\pi]$. The equal transmit power is allocated to each subcarrier, i.e., $P_k = P$, and the signal-to-noise ratio (SNR) is defined as $\mathrm{SNR} = \frac{P}{\sigma_n^2}$. The convergence threshold is set to $\epsilon = 10^{-3}$. In the following, we evaluate the performance achieved by the proposed schemes under different conditions.

A. Performance With Perfect CSI and Infinite Resolution Phase Shifters

In this subsection, we assume that perfect CSI and infinite resolution phase shifters are available. First of all, we evaluate the convergence property of the proposed AO-MM and AO-EBCD algorithms. It is worth noting that the convergence of iterative algorithms largely depends on the initialization of optimization variables. Therefore, we consider the following two initialization methods, namely, the proposed DFT-BD initialization (labeled with "DFT-BD ini") and random initialization (labeled with "Random ini").

Fig. 2 illustrates the spectral efficiency achieved by the proposed AO-MM and AO-EBCD algorithms versus the number of outer iterations I_{out}, where SNR is fixed at 0 dB. It is seen from Fig. 2 that the spectral efficiency of the proposed AO-MM and AO-EBCD algorithms with the DFT-BD initialization is monotonically increasing as the number of iterations increases and converges within a few iterations. This is due to the fact that the MM and EBCD methods are guaranteed to converge to stationary points. Nevertheless, the spectral efficiency of the proposed schemes with random initialization grows slowly and exhibits a significant performance gap compared to that with the DFT-BD initialization. The above results show that the proposed DFT-BD scheme can provide a high-quality initial point and speed up the convergence. Therefore, the DFT-BD scheme is utilized for the initialization of the AO-MM and AO-EBCD algorithms in the subsequent simulations.

Fig. 3 presents the spectral efficiency achieved by different schemes as a function of SNR. Besides, the sum capacity of MU-MIMO-OFDM systems attained by DPC [29] is also given to serve as the theoretical performance upper bound. We observe from Fig. 3 that the proposed AO-MM and AO-EBCD algorithms perform close to the fully digital WMMSE solution [31] and significantly outperform the existing hybrid transceiver designs [20], [21], [22]. Specifically, compared with the TD-RCD [22], AM-PE [21], and PCA-BD [20], the performance gains of the proposed AO-MM and AO-EBCD algorithms are over 1.5 dB, 3 dB, and 4 dB at the spectral efficiency of 30 bps/Hz, respectively. There are two main reasons for the performance improvement. On the one hand, different from the matrix factorization for approximating the fully digital solution [20], [21] and the two-stage processing for mitigating the multiuser interference [22], the proposed AO-MM and AO-EBCD algorithms directly tackle the spectral efficiency maximization problem. By exploiting the WMMSE equivalence and alternating optimization framework, the sophisticated original problem is decomposed into several tractable sub-problems, each of which is solved separately. On the other hand, the proposed AO-MM and AO-EBCD algorithms are guaranteed to converge to locally optimal solutions. Furthermore, it is also seen that the proposed DFT-BD scheme achieves comparable spectral efficiency to the state-of-the-art TD-RCD [22] and outperforms the other benchmarks [20], [21].

Fig. 4 shows the BER performance achieved by different schemes as a function of SNR, where 16-QAM symbols are adopted for data transmission at each subcarrier. It is observed from Fig. 4 that the proposed three hybrid transceiver designs exhibit significant performance improvement over the existing counterparts [20], [21], [22]. This is because the existing schemes only consider maximizing the beamforming gain in the analog domain, whereas the effect of the inter-beam interference is ignored. To address this issue, the proposed DFT-BD scheme selects the beam steering vectors from DFT codebook to form orthogonal beams, thus the inter-beam interference can be effectively mitigated. Benefited from the WMMSE criterion

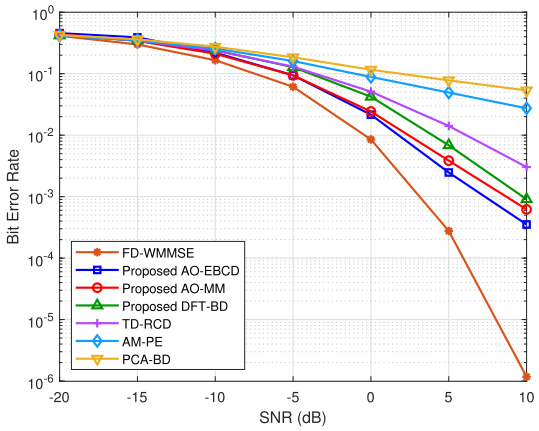

Fig. 4. BER versus SNR for different schemes using 16-QAM when $N_t = 64$, $N_r = 16$, $N_t^{\mathrm{RF}} = 8$, $N_r^{\mathrm{RF}} = 2$, $U = 4$, $N_s = 2$, and $K = 128$.

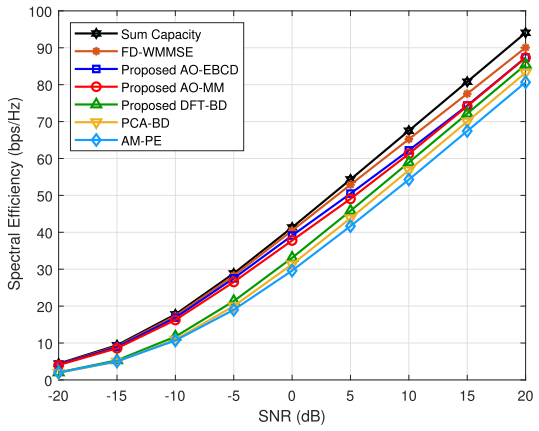

Fig. 5. Spectral efficiency versus SNR for different schemes when $N_t = 64$, $N_r = 16$, $N_t^{\mathrm{RF}} = 16$, $N_r^{\mathrm{RF}} = 4$, $U = 4$, $N_s = 2$, and $K = 128$.

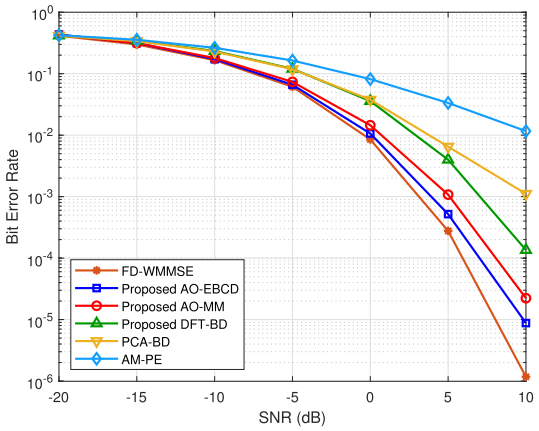

Fig. 6. BER versus SNR for different schemes using 16-QAM when $N_t = 64$, $N_r = 16$, $N_t^{\mathrm{RF}} = 16$, $N_r^{\mathrm{RF}} = 4$, $U = 4$, $N_s = 2$, and $K = 128$.

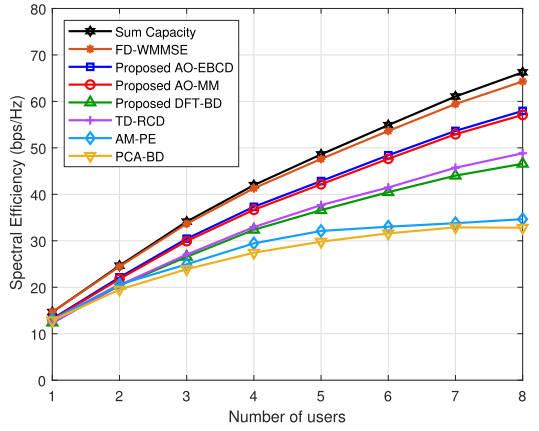

Fig. 7. Spectral efficiency versus the number of users U for different schemes when $N_t = 64$, $N_r = 16$, $N_r^{\mathrm{RF}} = 2$, $N_s = 2$, $K = 128$, and SNR $= 0$ dB.

and alternating optimization framework, the proposed AO-MM and AO-EBCD algorithms slightly outperform the proposed DFT-BD scheme.

Fig. 5 and Fig. 6 respectively present the spectral efficiency and BER performance achieved by different schemes versus SNR when the number of RF chains is twice the number of data streams, i.e., $N_t^{\mathrm{RF}} = 2UN_s$ and $N_r^{\mathrm{RF}} = 2N_s$. It is worth mentioning that the TD-RCD [22] is not included, since it is limited to the case where the number of RF chains is exactly equal to that of data streams. It is observed that the performance gains of the proposed AO-MM and AO-EBCD algorithms over the existing schemes [20], [21] further increase. Moreover, the proposed AO-MM and AO-EBCD algorithms can achieve similar performance to the fully digital WMMSE solution [31], since increasing the number of RF chains is capable of providing greater spatial degree of freedom.

Fig. 7 investigates the impact of the number of users U on the spectral efficiency achieved by different schemes, where $N_r^{\mathrm{RF}} = 2$, $N_t^{\mathrm{RF}} = UN_r^{\mathrm{RF}}$, and SNR is fixed at 0 dB. We observe from Fig. 7 that the performance gains of the proposed AO-MM and AO-EBCD algorithms over the state-of-the-art counterparts [20], [21], [22] become larger as the number of users increases. It is worth noting that when $U > 4$, the spectral efficiency of the proposed three schemes increases sustainably, whereas that of the PCA-BD [20] and the AM-PE [21] improves slowly and tends to be saturated. This is because most of existing schemes neglect the impact of inter-beam interference in the analog precoder design, which becomes more serious as the number of users increases. Different from these, the proposed DFT-BD scheme can effectively combat the inter-beam interference, thus enabling to accommodate more users.

Fig. 8 shows the spectral efficiency achieved by different schemes versus the number of transmit antennas N_t at the BS, where SNR is fixed at 0 dB. It is observed that the spectral efficiency of all schemes significantly improves with the increase of the number of antennas. This is due to the fact that more antennas can provide larger spatial degree of freedom and array gain. In addition, the proposed AO-MM and AO-EBCD algorithms are superior to the state-of-the-art hybrid transceiver schemes [20], [21], [22] and maintain almost constant performance gap with the fully digital WMMSE scheme [31].

B. Performance With Imperfect CSI

The aforementioned simulation results in subsection VI-A are based on the assumption that CSI is perfectly known. However, due to a large number of antennas and limited RF chains, it is extremely challenging to obtain accurate CSI for mmWave MU-MIMO-OFDM systems with hybrid

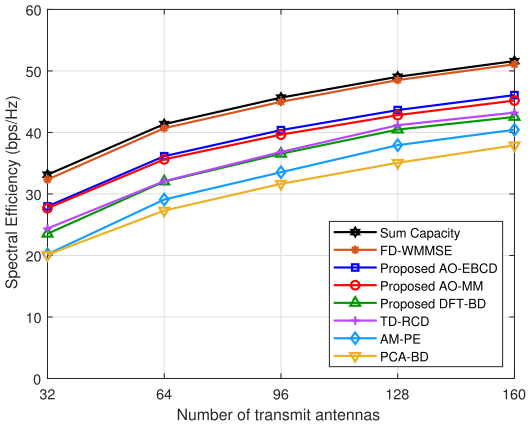

Fig. 8. Spectral efficiency versus the number of transmit antennas N_t for different schemes when $N_r = 16$, $N_t^{\mathrm{RF}} = 8$, $N_r^{\mathrm{RF}} = 2$, $U = 4$, $N_s = 2$, $K = 128$, and SNR $= 0$ dB.

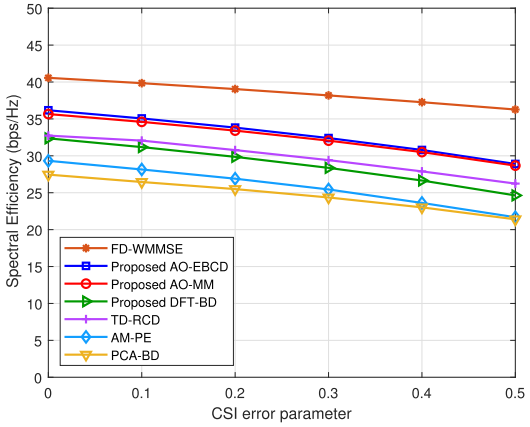

Fig. 9. Spectral efficiency versus the CSI error parameter δ for different schemes when $N_t = 64$, $N_r = 16$, $N_t^{\mathrm{RF}} = 8$, $N_r^{\mathrm{RF}} = 2$, $U = 4$, $N_s = 2$, $K = 128$, and SNR $= 0$ dB.

architecture. Therefore, we consider in this subsection that only imperfect CSI is available at the BS and all users and evaluate the robustness of the proposed schemes to channel estimation error. According to [42], the imperfect downlink channel between the BS to the u-th user at the k-th subcarrier can be expressed as

$$\hat{\mathbf{H}}_u[k] = \sqrt{1-\delta}\ \mathbf{H}_u[k] + \sqrt{\delta}\ \mathbf{H}_{e,u}[k], \tag{50}$$

where $\mathbf{H}_u[k]$ represents the perfect CSI, $\mathbf{H}_{e,u}[k]$ represents the channel estimation error with each element obeying $\mathcal{CN}(0,1)$, and δ is referred to as the CSI error parameter for characterizing the size of the CSI error. Clearly, $\delta = 0$ means that perfect CSI is available and $\delta = 1$ corresponds to that no CSI is available.

Fig. 9 presents the spectral efficiency achieved by different schemes as a function of the CSI error parameter δ, where SNR is fixed at 0 dB. We observe from Fig. 9 that the spectral efficiency of all schemes exhibit a downward trend as the channel estimation error increases, nevertheless, the proposed AO-MM and AO-EBCD still significantly outperform the existing hybrid transceiver schemes. Specifically, when $\delta = 0.5$, the performance loss of the proposed schemes does not exceed 20% relative to the perfect CSI case. Therefore, the proposed schemes are robust to channel estimation error.

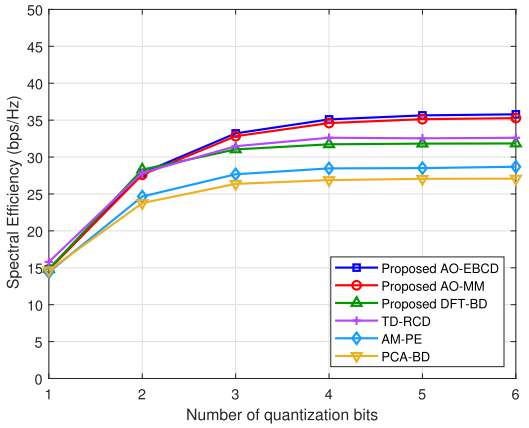

Fig. 10. Spectral efficiency versus the number of quantization bits for different schemes when $N_t = 64$, $N_r = 16$, $N_t^{\mathrm{RF}} = 8$, $N_r^{\mathrm{RF}} = 2$, $U = 4$, $N_s = 2$, $K = 128$, and SNR $= 0$ dB.

C. Performance With Finite Resolution Phase Shifters

In practical systems, low resolution phase shifters are generally applied to further reduce power consumption. In this subsection, we investigate the effect of the number of quantization bits of phase shifters on the performance achieved by the proposed schemes. For simplicity, the phase of the analog precoder and combiner is uniformly quantized.

Fig. 10 illustrates the spectral efficiency achieved by different schemes versus the number of quantization bits, where SNR is fixed at 0 dB. We observe that the proposed AO-MM and AO-EBCD schemes are superior to the existing hybrid transceiver designs. It is worth noting that, when the number of quantization bits exceeds 4, the spectral efficiency of all schemes tends to be saturated, thus the performance degradation caused by finite resolution phase shifters is negligible. In other words, only 4 quantization bits are required to achieve similar performance to infinite resolution phase shifters.

VII. Conclusion

This paper has investigated hybrid transceiver designs for mmWave massive MU-MIMO-OFDM systems. Firstly, we proposed two efficient algorithms based on alternating optimization for maximizing the spectral efficiency. Specifically, we transformed the intractable optimization problem into an equivalent one by exploiting the equivalence between the spectral efficiency maximization problem and the WMMSE problem. Then, the tractable problem was decoupled into several sub-problems with respect to each variable, which were solved separately. To design the analog precoder and combiner, we proposed an MM based algorithm and an EBCD based algorithm to handle the non-convex constant modulus constraint. The optimal digital precoder and combiner were obtained by applying the Lagrangian multiplier method and the first-order optimality condition, respectively. The proposed AO-MM and AO-EBCD algorithms were guaranteed to converge to locally optimal solutions. Moreover, to reduce the computational complexity, we proposed a DFT codebook based two-stage hybrid transceiver scheme, which effectively mitigated the inter-beam interference and provided a high-quality initial point for above two algorithms. Finally, simulation results demonstrated that the proposed AO-MM

and AO-EBCD algorithms achieved significant performance improvements over existing schemes under various conditions. Also, the proposed DFT-BD scheme achieved similar performance to state-of-the-art benchmark with reduced computational complexity.

It is worth pointing out that the proposed hybrid transceiver designs have strong scalability. The proposed AO-MM and AO-EBCD algorithms are easily extended to multi-cell [43] and cell-free [39] massive MIMO scenarios. For the future research directions, deep learning based hybrid transceiver design [44], joint active and passive beamforming design for reconfigurable intelligent surface assisted MIMO system [45], [46], [47], and blockage-aware beamforming for reliable mmWave communication [48] are also of great interests to be investigated.

APPENDIX
PROOF OF PROPOSITION 1

Proof: First, it is observed that the objective function of problem (11) is convex with respect to $\mathbf{W}_{\mathrm{BB},u}[k]$ and $\mathbf{\Lambda}_u[k]$, respectively. By checking the first-order optimality condition of problem (11), i.e.,

$$\frac{\partial \mathrm{Tr}(\mathbf{\Lambda}_u[k]\mathbf{E}_u[k])}{\partial(\mathbf{W}_{\mathrm{BB},u}[k])} = \mathbf{0}, \quad \forall k,u, \tag{51a}$$

$$\frac{\partial(\mathrm{Tr}(\mathbf{\Lambda}_u[k]\mathbf{E}_u[k]) - \log|\mathbf{\Lambda}_u[k]|)}{\partial(\mathbf{\Lambda}_u[k])} = \mathbf{0}, \quad \forall k,u, \tag{51b}$$

we can obtain the optimal digital combiner and weight matrix as

$$\mathbf{W}_{\mathrm{BB},u}[k] = \left(\mathbf{W}_{\mathrm{RF},u}^H\mathbf{H}_u[k]\left(\sum_{i=1}^{U}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB},i}[k]\mathbf{F}_{\mathrm{BB},i}^H[k]\mathbf{F}_{\mathrm{RF}}^H\right)\mathbf{H}_u^H[k]\mathbf{W}_{\mathrm{RF},u} + \sigma_n^2\mathbf{W}_{\mathrm{RF},u}^H\mathbf{W}_{\mathrm{RF},u}\right)^{-1} \times \mathbf{W}_{\mathrm{RF},u}^H\mathbf{H}_u[k]\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB},u}[k] \quad \forall k,u \tag{52a}$$

$$\mathbf{\Lambda}_u[k] = (\mathbf{E}_u[k])^{-1}, \quad \forall k,u. \tag{52b}$$

Further, by substituting the optimal digital combiner (52a) and weight matrix (52b) into the objective function of problem (11), we can obtain

$$\mathrm{Tr}(\mathbf{\Lambda}_u[k]\mathbf{E}_u[k]) - \log|\mathbf{\Lambda}_u[k]| - N_s \tag{53a}$$

$$= -\log\left|(\mathbf{E}_u[k])^{-1}\right| \tag{53b}$$

$$\overset{(a)}{=} -\log\left|\left(\mathbf{I}_{N_s} - \mathbf{F}_{\mathrm{BB},u}^H[k]\mathbf{F}_{\mathrm{RF}}^H\mathbf{H}_u^H[k]\mathbf{W}_{\mathrm{RF},u}(\mathbf{Q}_u[k])^{-1} \times \mathbf{W}_{\mathrm{RF},u}^H\mathbf{H}_u[k]\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB},u}[k]\right)^{-1}\right| \tag{53c}$$

$$\overset{(b)}{=} -\log\left|\mathbf{I}_{N_s} + \mathbf{F}_{\mathrm{BB},u}^H[k]\mathbf{F}_{\mathrm{RF}}^H\mathbf{H}_u^H[k]\mathbf{W}_{\mathrm{RF},u}(\mathbf{Y}_u[k])^{-1} \times \mathbf{W}_{\mathrm{RF},u}^H\mathbf{H}_u[k]\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB},u}[k]\right| \tag{53d}$$

$$\overset{(c)}{=} -\log\left|\mathbf{I}_{N_s} + (\mathbf{Y}_u[k])^{-1}\mathbf{W}_{\mathrm{RF},u}^H\mathbf{H}_u[k]\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB},u}[k] \times \mathbf{F}_{\mathrm{BB},u}^H[k]\mathbf{F}_{\mathrm{RF}}^H\mathbf{H}_u^H[k]\mathbf{W}_{\mathrm{RF},u}\right| \tag{53e}$$

$$\triangleq -\tilde{R}_{u,k} \tag{53f}$$

where we define

$$\mathbf{Q}_u[k] \triangleq \mathbf{W}_{\mathrm{RF},u}^H\mathbf{H}_u[k]\left(\sum_{i=1}^{U}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB},i}[k]\mathbf{F}_{\mathrm{BB},i}^H[k]\mathbf{F}_{\mathrm{RF}}^H\right) \times \mathbf{H}_u^H[k]\mathbf{W}_{\mathrm{RF},u} + \sigma_n^2\mathbf{W}_{\mathrm{RF},u}^H\mathbf{W}_{\mathrm{RF},u}, \tag{54a}$$

$$\mathbf{Y}_u[k] \triangleq \mathbf{W}_{\mathrm{RF},u}^H\mathbf{H}_u[k]\left(\sum_{i\neq u}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB},i}[k]\mathbf{F}_{\mathrm{BB},i}^H[k]\mathbf{F}_{\mathrm{RF}}^H\right) \times \mathbf{H}_u^H[k]\mathbf{W}_{\mathrm{RF},u} + \sigma_n^2\mathbf{W}_{\mathrm{RF},u}^H\mathbf{W}_{\mathrm{RF},u}, \tag{54b}$$

(a) is obtained by substituting (10) into (53b), (b) is derived by using the Woodbury matrix identity $(\mathbf{A}+\mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{C}^{-1}+\mathbf{DA}^{-1}\mathbf{B}\right)^{-1}\mathbf{DA}^{-1}$, (c) holds due to the fact that $|\mathbf{I}+\mathbf{AB}| = |\mathbf{I}+\mathbf{BA}|$.

Therefore, the WMMSE problem (11) can be reformulated as

$$\min_{\mathbf{F}_{\mathrm{RF}},\mathbf{F}_{\mathrm{BB}}[k],\mathbf{W}_{\mathrm{RF},u}} -\frac{1}{K}\sum_{k=1}^{K}\sum_{u=1}^{U}\tilde{R}_{u,k}$$
$$\text{s.t.} \quad \|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}[k]\|_F^2 \le P_k, \ \forall k,$$
$$\left|[\mathbf{F}_{\mathrm{RF}}]_{i,j}\right| = \frac{1}{\sqrt{N_t}}, \quad \forall i,j,$$
$$\left|[\mathbf{W}_{\mathrm{RF},u}]_{m,n}\right| = \frac{1}{\sqrt{N_r}}, \ \forall u,m,n. \tag{55}$$

Note that the spectral efficiency expression in (53e) does not include the digital combiner $\mathbf{W}_{\mathrm{BB},u}[k]$ explicitly compared with (7). Nevertheless, it is shown in [31] and [23] that the optimal MMSE digital combiner can achieve the maximum spectral efficiency. Therefore, the WMMSE problem (11) is equivalent to the original spectral efficiency maximization problem (9) and the globally optimal solutions of these two problems are identical. Thus, the proof is completed. ■

REFERENCES

[1] J. G. Andrews et al., "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[2] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 55–61, Mar. 2020.

[3] M. Xiao et al., "Millimeter wave communications for future mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1909–1935, Sep. 2017.

[4] S. He et al., "A survey of millimeter-wave communication: Physical-layer technology specifications and enabling transmission technologies," *Proc. IEEE*, vol. 109, no. 10, pp. 1666–1705, Oct. 2021.

[5] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[6] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: Ten myths and one critical question," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 114–123, Feb. 2016.

[7] R. W. Heath Jr., N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[8] E. Bjornson, L. Van der Perre, S. Buzzi, and E. G. Larsson, "Massive MIMO in sub-6 GHz and mmWave: Physical, practical, and use-case differences," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 100–108, Apr. 2019.

[9] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[10] S. Han, C.-L. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.

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

[12] J. Zhang, X. Yu, and K. B. Letaief, "Hybrid beamforming for 5G and beyond millimeter-wave systems: A holistic view," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 77–91, 2020.

[13] 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.

[14] 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.

[15] A. Alkhateeb, G. Leus, and R. W. Heath Jr., "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.

[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] M. Yuan, H. Wang, and Y. Sun, "BD-UCD-based nonlinear hybrid precoding for millimeter wave massive multiuser MIMO systems," *IEEE Commun. Lett.*, vol. 25, no. 3, pp. 1010–1014, Mar. 2021.

[18] 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.

[19] T. Lin, J. Cong, Y. Zhu, J. Zhang, and K. B. Letaief, "Hybrid beamforming for millimeter wave systems using the MMSE criterion," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3693–3708, May 2019.

[20] Y. Sun et al., "Principal component analysis-based broadband hybrid precoding for millimeter-wave massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6331–6346, Oct. 2020.

[21] H. Yuan, J. An, N. Yang, K. Yang, and T. Q. Duong, "Low complexity hybrid precoding for multiuser millimeter wave systems over frequency selective channels," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 983–987, Jan. 2019.

[22] G. Zilli and W.-P. Zhu, "Constrained tensor decomposition-based hybrid beamforming for mmWave massive MIMO-OFDM communication systems," *IEEE Trans. Veh. Technol.*, vol. 70, no. 6, pp. 5775–5788, Jun. 2021.

[23] Q. Shi and M. Hong, "Spectral efficiency optimization for millimeter wave multiuser MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 455–468, Jun. 2018.

[24] M. H. M. Costa, "Writing on dirty paper (corresp.)," *IEEE Trans. Inf. Theory*, vol. 29, no. 5, pp. 439–441, May 1983.

[25] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.

[26] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink–downlink duality," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.

[27] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[28] W. Yu and J. M. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1875–1892, Sep. 2004.

[29] N. Jindal, W. Rhee, S. Vishwanath, S. A. Jafar, and A. Goldsmith, "Sum power iterative water-filling for multi-antenna Gaussian broadcast channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1570–1580, Apr. 2005.

[30] S. S. Christensen, R. Agarwal, E. De Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.

[31] 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.

[32] A. Liao, Z. Gao, H. Wang, S. Chen, M.-S. Alouini, and H. Yin, "Closed-loop sparse channel estimation for wideband millimeter-wave full-dimensional MIMO systems," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8329–8345, Dec. 2019.

[33] X. Ma, Z. Gao, F. Gao, and M. Di Renzo, "Model-driven deep learning based channel estimation and feedback for millimeter-wave massive hybrid MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 8, pp. 2388–2406, Aug. 2021.

[34] M. Joham, W. Utschick, and J. A. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2700–2712, Aug. 2005.

[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] A. Arora, C. G. Tsinos, B. S. M. R. Rao, S. Chatzinotas, and B. Ottersten, "Hybrid transceivers design for large-scale antenna arrays using majorization-minimization algorithms," *IEEE Trans. Signal Process.*, vol. 68, pp. 701–714, 2020.

[37] S. Gong, C. Xing, V. K. N. Lau, S. Chen, and L. Hanzo, "Majorization-minimization aided hybrid transceivers for MIMO interference channels," *IEEE Trans. Signal Process.*, vol. 68, pp. 4903–4918, 2020.

[38] X. Zhai, X. Chen, J. Xu, and D. W. K. Ng, "Hybrid beamforming for massive MIMO over-the-air computation," *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2737–2751, Apr. 2021.

[39] C. Feng, W. Shen, J. An, and L. Hanzo, "Weighted sum rate maximization of the mmWave cell-free MIMO downlink relying on hybrid precoding," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2547–2560, Apr. 2022.

[40] 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.

[41] A. Alkhateeb and R. W. Heath Jr., "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801–1818, May 2016.

[42] S. Jacobsson, G. Durisi, M. Coldrey, and C. Studer, "Linear precoding with low-resolution DACs for massive MU-MIMO-OFDM downlink," *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 1595–1609, Mar. 2019.

[43] X. Chen, A. Liu, Y. Cai, V. K. N. Lau, and M. Zhao, "Randomized two-timescale hybrid precoding for downlink multicell massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 16, pp. 4152–4167, Aug. 2019.

[44] K. M. Attiah, F. Sohrabi, and W. Yu, "Deep learning for channel sensing and hybrid precoding in TDD massive MIMO OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 12, pp. 10839–10853, Dec. 2022.

[45] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, Nov. 2019.

[46] C. Pan et al., "Multicell MIMO communications relying on intelligent reflecting surfaces," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5218–5233, Aug. 2020.

[47] R. Li, B. Guo, M. Tao, Y.-F. Liu, and W. Yu, "Joint design of hybrid beamforming and reflection coefficients in RIS-aided mmWave MIMO systems," *IEEE Trans. Commun.*, vol. 70, no. 4, pp. 2404–2416, Apr. 2022.

[48] D. Kumar, J. Kaleva, and A. Tolli, "Blockage-aware reliable mmWave access via coordinated multi-point connectivity," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4238–4252, Jul. 2021.

Minghao Yuan received the B.S. degree in electronic information engineering from the China University of Geosciences, Beijing, China, in 2018, and the M.S. degree in information and communication engineering from the Beijing Institute of Technology, Beijing, in 2022, where he is currently pursuing the Ph.D. degree with the School of Information and Electronics. His current research interests include massive MIMO, millimeter-wave and terahertz communications, hybrid beamforming, and convex optimization.

Hua Wang (Member, IEEE) received the Ph.D. degree from the Beijing Institute of Technology, Beijing, China, in 1999. From February 2009 to January 2010, he was a Visiting Professor with the Department of Electrical Engineering, Arizona State University, Tempe, AZ, USA. He is currently a Professor with the School of Information and Electronics, Beijing Institute of Technology. His research interests include modulation and signal processing, wireless networking, and modem design and implementation for satellite communication.

Hao Yin received the B.S. degree in microwave communication and the M.S. degree in communication and information systems from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 1982 and 1987, respectively, and the Ph.D. degree in communication and information systems from the Beijing Institute of Technology, Beijing, China, in 1999. He is currently an Adjunct Professor with the Army Engineering University of PLA, Nanjing, and a Researcher with the Institute of China Electronic System Engineering Corporation, Beijing. His research interests include wireless communication networks and information systems. He is a fellow of the Chinese Academy of Sciences and the China Institute of Communications.

Dongxuan He received the B.S. degree in automation and the Ph.D. degree in information and communication systems from the Beijing Institute of Technology (BIT) in 2013 and 2019, respectively. From 2017 to 2018, he was a Visiting Student with the Singapore University of Technology and Design (SUTD). From 2019 to 2022, he was a Post-Doctoral Researcher with the Department of Electronic Engineering, Tsinghua University. He is currently an Assistant Professor with the School of Information and Electronics, BIT. His current research interests include terahertz communication, AI empowered wireless communications, and physical layer security. He was an Exemplary Reviewer of IEEE WIRELESS COMMUNICATIONS LETTERS.