IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 1, 2020, accepted June 12, 2020, date of publication June 16, 2020, date of current version June 30, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.3002788*

# Alternating Optimization Based Hybrid Precoding Strategies for Millimeter Wave MIMO Systems

**XU QIAO, YAO ZHANG, (Student Member, IEEE), MENG ZHOU, (Student Member, IEEE), AND LONGXIANG YANG**

Wireless Communications Key Laboratory of Jiangsu Province, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

Corresponding author: Longxiang Yang (yanglx@njupt.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFC1314903, in part by the National Natural Science Foundation of China under Grant 61861039, Grant 61372124, and Grant 61427801, in part by the Science and Technology Project Foundation of Gansu Province under Grant 18YF1GA060, and in part by the Postgraduate Research and Practice Innovation Program of Jiangsu Province under Grant SJKY19_0740.

**ABSTRACT** In millimeter wave (mmWave) multiple-input multiple-output (MIMO) systems, the hybrid beamforming architecture has been put forward to reduce the high hardware cost and power consumption, which are resulted from the tremendous requirements of dedicated radio frequency (RF) chains. In this paper, we propose several strategies to design analog and digital precoders for a point-to-point (P2P) hybrid MIMO system. Aiming at minimizing the Euclidean distance between the optimal digital precoder and hybrid precoder, we decouple this matrix factorization problem into a nonconvex quadratically constrained quadratic programming (QCQP) problem and an unit-modulus least-squares (ULS) problem, which can be solved by the presented three alternating optimization algorithms. Simulation and analysis results indicate that the proposed semidefinite relaxation based alternating optimization (SDR−AO) algorithm can approach near-optimal spectral efficiency performance compared with previous algorithms in the literature, but shows extremely high computational complexity. The alternating direction method of multipliers based alternating optimization (ADMM−AO) algorithm is preferred in the case that the number of transmit antennas is much larger than that of receive antennas or the amount of data streams is small. Moreover, when equal number of RF chains and data streams are employed, the analytical constant modulus factorization based alternating optimization (ACMF−AO) algorithm is a better choice. Finally, the proposed algorithms can also be well applied in finite resolution phase shifters (PSs) of the analog component and are extended to wideband mmWave systems.

**INDEX TERMS** Millimeter wave communication, hybrid precoding, alternating optimization, quadratically constrained quadratic programming.

## I. INTRODUCTION

Millimeter wave (mmWave) wireless communication has proven to be one of the prime candidates for next-generation cellular systems offering large bandwidth and high data rates [1], [2]. Combined with preprocessing and post-processing techniques in multiple-input multiple-output (MIMO) systems, the large antenna array at transceivers will provide sufficient beamforming gains to combat the severe path loss in mmWave channel [3]–[5]. However, new hardware limitations and large-scale antennas will bring new challenges. In traditional low-frequency (below 6 GHz) MIMO systems, the transmit signals are adjusted by baseband digital precoders, then connected to each antenna after passing though the radio frequency (RF) chain. In this configuration, the number of RF chains must be equal to the number of antenna elements. However, deploying such a dedicated RF chain for each antenna in mmWave (30−300 GHz) MIMO systems will result in prohibitively high hardware cost and power consumption.

To overcome this challenge, a hybrid transceiver architecture has recently been put forward, in which full digital beamformers were replaced by analog/digital precoders and combiners. Benefiting from the use of analog precoders and combiners, typically implemented by phase shifters (PSs), the number of RF chains in hybrid MIMO architectures was greatly reduced. Besides, some economical and energy-efficiency devices such as low-resolution analog to

The associate editor coordinating the review of this manuscript and approving it for publication was Yiming Huo.

digital converters (ADCs) and digital to analog converters (DACs) can also be adopted, since the power consumption of ADC/DAC grows exponentially with the quantization bit [6]–[8].

The pioneering work on hybrid precoding and combining design for mmWave channel was [9]. Leveraging on the spatial structure of mmWave channel, the design of analog and digital precoder/combiner were formulated as a sparse signal recovery problem and solved by the orthogonal matching pursuit (OMP) algorithm. After that, aiming at solving the matrix factorization problem with constant modulus constraints presented in [9], several alternating minimization (AltMin) algorithms were developed [10]–[13]. The phase extraction based AltMin (PE−AltMin) algorithm and the iterative coordinate descent algorithm (ICDA) were proposed in [10] and [11], respectively. The authors in [12] decoupled the nonconvex matrix decomposition problem into a series of convex subproblems. Subsequently, four hybrid design algorithms were developed in [13] and the preferred algorithm depended on the implementation complexity, power consumption, and other system characteristics. In addition, it is also a good trick to minimize the Euclidean distance between hybrid precoding matrix and full-digital precoding matrix by relaxing the objective function or constraints appropriately [14], [15].

Furthermore, in extended multi-user scenarios, the analog precoder/combiner were designed by maximizing the user's signal power. Then, based on zero forcing (ZF) [16], block diagonalization (BD) [17], and minimum mean square error (MMSE) [18], [19] processing methods, the digital beamformer were devised to eliminate inter-user interference (IUI). Recently, hybrid beamforming with sub-connected structure [10], [20]–[22], dynamic sub-connected structure [23], [24] and distributed architecture [25]–[27] have been investigated. In order to further reducing the hardware complexity, the switching networks [28] and the low-resolution PSs [29] can be employed to further reduce the power consumption caused by analog beamformer. However, the reduction of hardware complexity is at the expense of performance loss.

The aforementioned algorithms for solving matrix decomposition problems with constant modulus constraint make certain assumptions about the analog precoders or digital precoders without exception. For instance, the authors in [9] restrict the feasible solution of analog precoding to the array response vectors of mmWave channel matrix. Moreover, the extra orthgonal property was imposed in digital precoder in [10]. Although these assumptions simplify the design of the hybrid precoding matrix, it inevitably results in performance degradation of spectral efficiency.

In this paper, starting from minimizing the Euclidean distance between the optimal unconstrained digital precoding matrix and the products of analog RF and digital baseband precoding matrices, subjecting to constant modulus constraint and power constraint, we propose three alternating optimization strategies based on different principles to design hybrid precoding matrix. The proposed methods are also applicable to the design of hybrid combiner at receiver due to the similar structure and make no assumptions about the properties of the analog/digital precoder. Moreover, the presented algorithms can be extended for OFDM-based wideband mmWave MIMO systems. The computational complexity analyses and simulation results demonstrate that the proposed semidefinite relaxation based alternating optimization (SDR−AO) algorithm can achieve near-optimal spectral efficiency performance compared with other state-of-the-art algorithm in the literature, but at cost of extremely high computational complexity. Although the complexity of presented the alternating direction method of multipliers based alternating optimization (ADMM−AO) and analytical constant modulus factorization based alternating optimization (ACMF−AO) algorithms are one order of magnitude higher than OMP and ICDA algorithm, the spectral efficiencies of these two presented algorithms are far superior to them. More specifically, ADMM−AO is preferred in the case that the number of transmit antennas is much larger than that of receive antennas or the amount of data streams is small. Furthermore, when equal number of RF chains and data streams are employed, ACMF−AO is a better choice since it achieves a trade-off between computational complexity and spectral efficiency performance. Finally, the simulation results also show that the proposed algorithms can be greatly applied in uniform quantized PSs of the analog component.

The remainder of this paper is organized as follows. Section II introduces the data transmission model and channel model. Section III presents the formulation of optimization problem and the proposed alternating optimization algorithms. In Section IV, the proposed algorithms are extended for mmWave wideband systems. The computational complexity analyses and simulation results are provided in Section V. Finally, conclusions are drawn in Section VI.

## II. SYSTEM MODEL

### A. DATA TRANSMISSION MODEL

This paper focuses on a point-to-point (P2P) MIMO system with hybrid architecture as shown in [9]–[13]. The transmitter equipped with $N_{\mathrm{t}}$ antennas communicates $N_{\mathrm{s}}$ data streams with the receiver equipped with $N_{\mathrm{r}}$ antennas. The hybrid precoding and combining architecture are employed to substitute full digital beamforming architecture for reducing the deployment of RF chains. Between baseband precoder (combiner) and PSs, $L_{\mathrm{t}}$ ($L_{\mathrm{r}}$) RF chains are deployed to up-convert the complex transmit symbols to the passband domain, which are subjected to constraints $N_{\mathrm{s}} \le L_{\mathrm{t}} \le N_{\mathrm{t}}$ and $N_{\mathrm{s}} \le L_{\mathrm{r}} \le N_{\mathrm{r}}$ to enable multi-stream communication.

Assuming a narrowband flat fading channel and perfect synchronization, the $N_{\mathrm{s}} \times 1$ received signal vector after combiner processing can be written as

$$\mathbf{r} = \sqrt{P}\mathbf{W}_{\mathrm{B}}^{H}\mathbf{W}_{\mathrm{R}}^{H}\mathbf{H}\mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{B}}\mathbf{s} + \mathbf{W}_{\mathrm{B}}^{H}\mathbf{W}_{\mathrm{R}}^{H}\mathbf{n}, \tag{1}$$

where $P$ stands for the average received power, $\mathbf{W}_{\mathrm{B}} \in \mathbb{C}^{L_{\mathrm{r}} \times N_{\mathrm{s}}}$ means the digital baseband combiner, $\mathbf{W}_{\mathrm{R}} \in \mathbb{C}^{N_{\mathrm{r}} \times L_{\mathrm{r}}}$ is the analog RF combiner, and $\mathbf{H} \in \mathbb{C}^{N_{\mathrm{r}} \times N_{\mathrm{t}}}$ represents the complex

channel matrix. $\mathbf{F}_{\mathrm{R}} \in \mathbb{C}^{N_{\mathrm{t}} \times L_{\mathrm{t}}}$ and $\mathbf{F}_{\mathrm{B}} \in \mathbb{C}^{L_{\mathrm{t}} \times N_{\mathrm{s}}}$ denote the analog RF precoder and digital baseband precoder respectively, and $\mathbf{s}$ denotes the $N_{\mathrm{s}} \times 1$ vector of data symbols such that $\mathbb{E}\left\{\mathbf{s}\mathbf{s}^H\right\} = \frac{1}{N_{\mathrm{s}}}\mathbf{I}_{N_{\mathrm{s}}}$. Besides, $\mathbf{n}$ refers to the $N_{\mathrm{r}} \times 1$ vector of independent and identically distributed (*i.i.d.*) $\mathcal{CN}(0, \sigma_{\mathrm{n}}^2)$ additive white complex Gaussian noise.

In order to normalize the total power of the discrete-time transmitted signal which can be written as $\mathbf{x} = \mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{B}}\mathbf{s}$, the hybrid precoding matrix is enforced to $\|\mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{B}}\|_F^2 = N_{\mathrm{s}}$. Furthermore, since the RF precoders (combiners) are implemented by PSs, the hardware limitation will be satisfied, namely, the magnitude of all elements in $\mathbf{F}_{\mathrm{R}}$ ($\mathbf{W}_{\mathrm{R}}$) is constant. Therefore, $\mathbf{F}_{\mathrm{R}}(i,j) \in \mathcal{F}$ ($\mathbf{W}_{\mathrm{R}}(i,j) \in \mathcal{W}$), $\forall i, j$, in which $\mathcal{F}$ ($\mathcal{W}$) refers to the feasible set of RF precoders (combiners). For the infinite-resolution PSs, this hardware limitation is equivalent to the unit-modulus constraint $|\mathbf{F}_{\mathrm{R}}(i,j)|^2 = 1$ ($|\mathbf{W}_{\mathrm{R}}(i,j)|^2 = 1$). However, if $B$-bit uniform quantized PSs are considered, the feasible set can be defined by $\mathcal{F} \triangleq \left\{e^{j2\pi b/2^B} | b = 1, \ldots, 2^B\right\}$. In the following discussion, we first analyze the design of hybrid precoding with the assumption of infinite resolution PSs. Afterward, in the case of $B$-bit resolution PSs, we capture $\mathbf{F}_{\mathrm{R}}$ by simply selecting the entries of $\mathcal{F}$ that are closest to the RF precoder gained in infinite resolution PSs.

In this paper, perfect channel state information (CSI) is assumed to known at both transmitter and receiver. When Gaussian symbols are employed to transmit over the channel, the spectral efficiency shall be

$$R = \log_2 \left| \mathbf{I}_{N_{\mathrm{s}}} + \frac{P}{\sigma_{\mathrm{n}}^2 N_{\mathrm{s}}} \mathbf{T}_{\mathrm{w}}^{-1} \mathbf{W}_{\mathrm{B}}^H \mathbf{W}_{\mathrm{R}}^H \mathbf{H} \mathbf{F}_{\mathrm{R}} \mathbf{F}_{\mathrm{B}} \times \mathbf{F}_{\mathrm{B}}^H \mathbf{F}_{\mathrm{R}}^H \mathbf{H}^H \mathbf{W}_{\mathrm{R}} \mathbf{W}_{\mathrm{B}} \right|, \quad (2)$$

where $\mathbf{T}_{\mathrm{w}} = \mathbf{W}_{\mathrm{B}}^H \mathbf{W}_{\mathrm{R}}^H \mathbf{W}_{\mathrm{R}} \mathbf{W}_{\mathrm{B}}$. For the sake of simplifying the hybrid beamformer design, we temporarily focus on the design of RF and baseband precoder $\mathbf{F}_{\mathrm{R}}$, $\mathbf{F}_{\mathrm{B}}$ assuming that prefect hybrid combiner are employed in receiver. Hence, the spectral efficiency without considering combiner processing can be expressed as

$$R = \log_2 \left| \mathbf{I}_{N_{\mathrm{r}}} + \frac{P}{\sigma_{\mathrm{n}}^2 N_{\mathrm{s}}} \mathbf{H} \mathbf{F}_{\mathrm{R}} \mathbf{F}_{\mathrm{B}} \mathbf{F}_{\mathrm{B}}^H \mathbf{F}_{\mathrm{R}}^H \mathbf{H}^H \right|. \quad (3)$$

It is worth noting that the structure of hybrid combiner and precoder are similar, then the next proposed alternating optimization algorithm for designing hybrid precoder can also be employed to the design of hybrid combiner. In the next section, we will focus on the design of hybrid analog and digital precoding algorithm.

### B. CHANNEL MODEL

The limited scattering characteristic of mmWave frequency band will result in a sparse channel structure. Thus, in this paper, the clustered channel model based on extended Saleh-Valenzuela model is adopted, which allows us to accurately capture the mathematical property present in mmWave channels. The discrete-time narrowband channel matrix $\mathbf{H}$ [9]–[13] is given by

$$\mathbf{H} = \sqrt{\frac{N_{\mathrm{t}} N_{\mathrm{r}}}{N_{\mathrm{c}} N_{\mathrm{p}}}} \sum_{i=1}^{N_{\mathrm{c}}} \sum_{j=1}^{N_{\mathrm{p}}} \beta_{i,j} \mathbf{a}_{\mathrm{r}}(\phi_{i,j}) \mathbf{a}_{\mathrm{t}}(\theta_{i,j})^H, \quad (4)$$

where $N_{\mathrm{c}}$ stands for the number of scattering clusters, each of which include $N_{\mathrm{p}}$ propagation paths. $\beta_{i,j}$ is the complex gain of the $j^{\mathrm{th}}$ propagation ray in the $i^{\mathrm{th}}$ scattering cluster, which is assumed following Gaussian distribution $\mathcal{CN}(0, \sigma_{\alpha,i}^2)$. $\mathbf{a}_{\mathrm{r}}(\phi_{i,j})$ and $\mathbf{a}_{\mathrm{t}}(\theta_{i,j})$ represent the normalized receive and transmit antenna array response vectors at the angle of $\phi_{i,j}$ and $\theta_{i,j}$, respectively, while $\phi_{i,j}$ and $\theta_{i,j}$ denote the azimuth angles of arrival and departure (AoA and AoD), respectively.

Since array geometry will affect the dictionary matrix, here, we choose the case of uniform linear array (ULA) with $N$ elements to form the dictionary matrix. The array response vector at angle of $\theta$ has the form

$$\mathbf{a}_{\mathrm{ULA}}(\theta) = \frac{1}{\sqrt{N}} \left[1, e^{jkd\sin(\theta)}, \ldots, e^{jkd(N-1)\sin(\theta)}\right]^T, \quad (5)$$

where $k = 2\pi/\lambda$ and $\lambda$ is the carrier wavelength, $d$ denotes the distance between two adjacent antenna elements. The array response vectors at both the transmitter and the receiver can be written in the form of (5). Although this channel model will be adopted in our simulation, the following proposed hybrid precoding algorithm is not only just limited to the above mmWave channel, but also can be applied to more general models.

## III. PROPOSED ALTERNATING OPTIMIZATION ALGORITHMS

### A. THE FORMULATION OF OPTIMIZATION PROBLEM

As shown in reference [9], the hybrid precoders design problem concerned with maximize spectral efficiency (3) is equivalent to minimize the Euclidean distance between optimal unconstrained digital precoder and hybrid precoders, i.e. $\left\|\mathbf{F}_{\mathrm{opt}} - \mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{B}}\right\|_F$. By imposing the limitation on the RF and baseband precoder, the corresponding design problem can be written in the following form

$$(\mathcal{P}_1): \min_{\mathbf{F}_{\mathrm{R}}, \mathbf{F}_{\mathrm{B}}} \left\|\mathbf{F}_{\mathrm{opt}} - \mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{B}}\right\|_F^2 \quad (6a)$$

$$\text{s.t. } \mathbf{F}_{\mathrm{R}}(i,j) \in \mathcal{F}, \ \forall i, j, \quad (6b)$$

$$\|\mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{B}}\|_F^2 = N_{\mathrm{s}}, \quad (6c)$$

where $\mathbf{F}_{\mathrm{opt}} \in \mathbb{C}^{N_{\mathrm{t}} \times N_{\mathrm{s}}}$ stands for optimal unconstrained digital precoding matrix, which consists of the first $N_{\mathrm{s}}$ right singular vectors of $\mathbf{H}$. Therefore, we have $\mathbf{F}_{\mathrm{opt}} = \mathbf{V}_1$, where $\mathbf{V}_1$ derived from the singular value decomposition (SVD) of channel matrix, i.e. $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$. Notice that the singular values in $\mathbf{\Sigma}$ are all non-negative real numbers and sorted in descending order. Similar to the structure of $(\mathcal{P}_1)$, the hybrid combining problem can also be formulated as

$$\min_{\mathbf{W}_{\mathrm{R}}, \mathbf{W}_{\mathrm{B}}} \left\|\mathbf{W}_{\mathrm{opt}} - \mathbf{W}_{\mathrm{R}}\mathbf{W}_{\mathrm{B}}\right\|_F^2 \quad (7a)$$

$$\text{s.t. } \mathbf{W}_{\mathrm{R}}(i,j) \in \mathcal{F}, \ \forall i, j. \quad (7b)$$

Notably, the above hybrid combining problem is akin to $(\mathcal{P}_1)$ except for an extra power constraint (6c). Thus, we pay attention to $(\mathcal{P}_1)$ because the developed algorithm can also be applied to the design of hybrid combining. Unfortunately, the problem $(\mathcal{P}_1)$ is nonconvex due to the hardware constraint (6b) and power constraint (6c), additionally, optimizing $\mathbf{F}_{\mathrm{R}}$, $\mathbf{F}_{\mathrm{B}}$ simultaneously make the problem intractable. In this paper, $(\mathcal{P}_1)$ is viewed as a matrix factorization problem and the alternating optimization strategies will be adopted to deal with it. Based on the principle of alternating optimization, we will alternately optimize matrix $\mathbf{F}_{\mathrm{R}}$ or $\mathbf{F}_{\mathrm{B}}$ while fixing the other. In this case, the original problem is decoupled into two unrelated optimization subproblems. Assuming that the RF precoding matrix $\mathbf{F}_{\mathrm{R}}$ is given, $(\mathcal{P}_1)$ will be simplified as

$$(\mathcal{P}_2): \min_{\mathbf{F}_{\mathrm{B}}} \ \left\|\mathbf{F}_{\mathrm{opt}}-\mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{B}}\right\|_F^2 \tag{8a}$$

$$\text{s.t.} \ \ \|\mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{B}}\|_F^2 = N_{\mathrm{s}}. \tag{8b}$$

Obviously, the problem $(\mathcal{P}_2)$ is a nonconvex quadratically constrained quadratic programming (QCQP) substantially [30], which can be reformulated as a semidefinite programming (SDP) by relaxation. We will analyze the problem elaborately in the following subsection III-B. After solving $(\mathcal{P}_2)$, we will update $\mathbf{F}_{\mathrm{B}}$ by fixing $\mathbf{F}_{\mathrm{R}}$ and the RF precoding matrix design problem will be accordingly formulated as

$$(\mathcal{P}_3): \min_{\mathbf{F}_{\mathrm{R}}} \ \left\|\mathbf{F}_{\mathrm{opt}}-\mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{B}}\right\|_F^2 \tag{9a}$$

$$\text{s.t.} \ \ |\mathbf{F}_{\mathrm{R}}(i,j)|^2 = 1, \quad \forall i,j. \tag{9b}$$

The problem $(\mathcal{P}_3)$ will equal to an unit-modulus least-squares (ULS) problem after matrix vectorization, and some efficient algorithms will be developed to solve it in next. We will solve $(\mathcal{P}_2)$ and $(\mathcal{P}_3)$ alternately and finally obtain the hybrid precoders. On top of this, solving $(\mathcal{P}_2)$ always happen after solving $(\mathcal{P}_3)$ to ensure the normalized transmitting power.

### *B. SDR–AO: SEMIDEFINITE RELAXATION BASED ALTERNATING OPTIMIZATION*

We first execute vectorization on $(\mathcal{P}_3)$ to transform the original problem into an ULS problem [31]. The objective function in (9a) can be rewritten as

$$\begin{aligned}&\left\|\mathbf{F}_{\mathrm{opt}}-\mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{B}}\right\|_F^2\\ &\quad=\left\|\mathrm{vec}\left(\mathbf{F}_{\mathrm{opt}}-\mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{B}}\right)\right\|_2^2\\ &\quad=\left\|\mathrm{vec}\left(\mathbf{F}_{\mathrm{opt}}\right)-\left(\mathbf{F}_{\mathrm{B}}^T\otimes\mathbf{I}_{N_{\mathrm{t}}}\right)\mathrm{vec}\left(\mathbf{F}_{\mathrm{R}}\right)\right\|_2^2.\end{aligned} \tag{10}$$

To make the notation concise, we denote $\mathbf{y}=\mathrm{vec}\left(\mathbf{F}_{\mathrm{opt}}\right)$, $\mathbf{A}=\mathbf{F}_{\mathrm{B}}^T\otimes\mathbf{I}_{N_{\mathrm{t}}}$, and $\mathbf{w}=\mathrm{vec}\left(\mathbf{F}_{\mathrm{R}}\right)$. Therefore, $(\mathcal{P}_3)$ will be restated by

$$(\mathcal{P}_4): \min_{\mathbf{w}} \ \|\mathbf{y}-\mathbf{A}\mathbf{w}\|_2^2 \tag{11a}$$

$$\text{s.t.} \ \ |\mathbf{w}_i|^2 = 1, \quad i=1,\ldots,N_{\mathrm{t}}L_{\mathrm{t}}. \tag{11b}$$

With the introduction of extra auxiliary variable $t\in\mathbb{C}$, the above ULS problem $(\mathcal{P}_4)$ can be further homogenized to an unit-modulus quadratic programming (UQP) [32]

$$(\mathcal{P}_5): \min_{\bar{\mathbf{w}}} \ \bar{\mathbf{w}}^H\mathbf{R}\bar{\mathbf{w}} \tag{12a}$$

$$\text{s.t.} \ \ |\bar{\mathbf{w}}_i|^2 = 1, \quad i=1,\ldots,N_{\mathrm{t}}L_{\mathrm{t}}+1, \tag{12b}$$

where

$$\mathbf{R}=\begin{bmatrix}\mathbf{A}^H\mathbf{A} & -\mathbf{A}^H\mathbf{y}\\ -\mathbf{y}^H\mathbf{A} & \mathbf{y}^H\mathbf{y}\end{bmatrix}, \quad \bar{\mathbf{w}}=\begin{bmatrix}\mathbf{w}\\ t\end{bmatrix}. \tag{13}$$

Notice that problem $(\mathcal{P}_4)$ is equivalent to $(\mathcal{P}_5)$ in the following sense: $[(\mathbf{w}^\star)^T, t^\star]^T$ is an optimal solution of $(\mathcal{P}_5)$ if and only if $\mathbf{w}^\star$ is an optimal solution of $(\mathcal{P}_4)$. To deal with the UQP $(\mathcal{P}_5)$, an classical processing method is semedefinite relaxation (SDR) [33]. Rewriting the objective function in (12a) as $\bar{\mathbf{w}}^H\mathbf{R}\bar{\mathbf{w}}=\mathrm{Tr}(\bar{\mathbf{w}}^H\mathbf{R}\bar{\mathbf{w}})=\mathrm{Tr}(\mathbf{R}\bar{\mathbf{w}}\bar{\mathbf{w}}^H)$ and defining an matrix $\mathbf{W}$ as $\mathbf{W}\triangleq\bar{\mathbf{w}}\bar{\mathbf{w}}^H$, $(\mathcal{P}_5)$ can be equivalently reformulated by the following problem

$$\min_{\mathbf{W}\in\mathbb{H}^n} \ \mathrm{Tr}\left(\mathbf{R}\mathbf{W}\right) \tag{14a}$$

$$\text{s.t.} \ \ \mathbf{W}(i,i)=1, \quad i=1,\ldots,n, \tag{14b}$$

$$\mathbf{W}\succeq 0, \tag{14c}$$

$$\mathrm{rank}(\mathbf{W})=1, \tag{14d}$$

where $\mathbb{H}^n$ denotes the set of $n\times n$ ($n=N_{\mathrm{t}}L_{\mathrm{t}}+1$) Hermitian matrices, the constraint $\mathbf{W}\succeq 0$ means that $\mathbf{W}$ is a positive semidefinite matrix [34]. Except for the rank constraint, the other constraints and objective functions in (14) are convex. Thus, we will temporarily ignore the rank-1 constraint and yield a relaxed form of (14) which is convex:

$$(\mathcal{P}_6): \min_{\mathbf{W}\in\mathbb{H}^n} \ \mathrm{Tr}\left(\mathbf{R}\mathbf{W}\right) \tag{15a}$$

$$\text{s.t.} \ \ \mathbf{W}(i,i)=1, \quad i=1,\ldots,n, \tag{15b}$$

$$\mathbf{W}\succeq 0. \tag{15c}$$

**Algorithm 1** SDR−AO Based Hybrid Precoding Design

**Input**: $\mathbf{F}_{\mathrm{opt}}$

1: Initialize $\mathbf{F}_{\mathrm{R}}^{(0)}$ with random phases;
2: $\mathbf{F}_{\mathrm{B}}^{(0)}=(\mathbf{F}_{\mathrm{R}}^{(0)})^{+}\mathbf{F}_{\mathrm{opt}}$;
3: $\mathbf{F}_{\mathrm{B}}^{(0)}=\sqrt{N_{\mathrm{s}}}/\|\mathbf{F}_{\mathrm{R}}^{(0)}\mathbf{F}_{\mathrm{B}}^{(0)}\|_F\cdot\mathbf{F}_{\mathrm{B}}^{(0)}$;
4: $k=0$;
5: **repeat**
6: Fix $\mathbf{F}_{\mathrm{B}}^{(k)}$, obtain $\mathbf{W}$ by solving $(\mathcal{P}_6)$ with CVX toolbox;
7: Update $\mathbf{F}_{\mathrm{R}}^{(k+1)}$ using rank-1 approximation and matrix reconstruction;
8: Fix $\mathbf{F}_{\mathrm{R}}^{(k+1)}$, obtain $\mathbf{Z}$ by solving (16) with CVX toolbox;
9: Update $\mathbf{F}_{\mathrm{B}}^{(k+1)}$ using rank-1 approximation and matrix reconstruction;
10: $k=k+1$;
11: **until** a stopping criterion triggers

**Output**: $\mathbf{F}_{\mathrm{R}}, \mathbf{F}_{\mathrm{B}}$

The problem $(\mathcal{P}_6)$ is well-known as a SDP, which can be solved readily and effectively by convex optimization toolbox CVX [30], [33]. However, the optimal solution $\mathbf{W}^\star$ of $(\mathcal{P}_6)$ is usually not rank one. Then, to obtain the feasible solution $\bar{\mathbf{w}}$, some efficient heuristic methods such as randomization [33], [35] will be exploited to extract from $\mathbf{W}^\star$. In this paper, we utilize a simple rank-1 approximation on $\mathbf{W}^\star$, i.e. $\mathbf{W}_1^\star = \lambda_1 \mathbf{q}_1 \mathbf{q}_1^H$, where $\lambda_1$ is the largest eigenvalue and $\mathbf{q}_1$ is the eigenvector associated with $\lambda_1$. After that, an near-optimal solution of $(\mathcal{P}_5)$ can be written by $\bar{\mathbf{w}} = \sqrt{\lambda_1}\mathbf{q}_1$ (it is in general not an optimal solution unless $\text{rank}(\mathbf{W}^\star) = 1$). At last, removing the last entry of $\bar{\mathbf{w}}$ to form $\mathbf{w}$ and the RF precoding matrix will be reshaped by $\mathbf{F}_\text{R} = \text{vec}^{-1}(\mathbf{w})$ eventually.[1] So far, the design of RF precoder has been completed and we will continue to solve $(\mathcal{P}_2)$ to get baseband precoder.

As mentioned before, the nonconvex QCQP $(\mathcal{P}_2)$ can be transformed into a SDP. Similar to the vectorization and homogenization procedures in $(\mathcal{P}_4)$, $(\mathcal{P}_5)$ and leveraging on the same SDR method (the rank-1 constraint is also dropped), $(\mathcal{P}_2)$ will be also reformulated as

$$\min_{\mathbf{Z}\in\mathbb{H}^m} \text{Tr}(\mathbf{CZ}) \tag{16a}$$
$$\text{s.t. } \text{Tr}\,(\mathbf{Q}_1\mathbf{Z}) = N_\text{s}, \tag{16b}$$
$$\text{Tr}\,(\mathbf{Q}_2\mathbf{Z}) = 1, \tag{16c}$$
$$\mathbf{Z} \succeq 0, \tag{16d}$$

where $m = L_\text{t}N_\text{s} + 1$ and $\mathbf{Z} = \bar{\mathbf{z}}\bar{\mathbf{z}}^H$ is a $m \times m$ Hermitian matrix. Besides, $\bar{\mathbf{z}} = [\mathbf{z}^T, t]^T$, $\mathbf{z} = \text{vec}^T(\mathbf{F}_\text{B})$, denoting $\mathbf{B} = \mathbf{I}_{N_\text{s}} \otimes \mathbf{F}_\text{R}$ and $\mathbf{C}$, $\mathbf{Q}_1$, $\mathbf{Q}_2$ is defined by

$$\mathbf{C} = \begin{bmatrix} \mathbf{B}^H\mathbf{B} & -\mathbf{B}^H\mathbf{y} \\ -\mathbf{y}^H\mathbf{B} & \mathbf{y}^H\mathbf{y} \end{bmatrix}, \quad \mathbf{Q}_1 = \begin{bmatrix} \mathbf{B}^H\mathbf{B} & 0 \\ 0 & 0 \end{bmatrix},$$
$$\mathbf{Q}_2 = \begin{bmatrix} 0_{(m-1)\times(m-1)} & 0 \\ 0 & 1 \end{bmatrix}. \tag{17}$$

The SDP (16) can also be solved by the CVX toolbox, and the baseband precoding matrix is harvested by matrix reconstruction $\mathbf{F}_\text{B} = \text{vec}^{-1}(\mathbf{z})$ after the rank-1 approximation. To sum up, the overall proposed procedure for designing the hybrid precoders via SDR−AO is given in **Algorithm 1**.

In addition to the above SDR method for designing $\mathbf{F}_\text{B}$, a simple scaling method is available in [10], [13], [15] to achieve the closed-form solution of baseband precoding matrix. Firstly, without accounting for the power constraint (8b), $(\mathcal{P}_2)$ is simplified as an unconstrained least squares (LS) problem, we can easily get the solution that is $\mathbf{F}_\text{B} = \mathbf{F}_\text{R}^+\mathbf{F}_\text{opt}$. After that, the LS solution is multiplied by an scaling factor $\gamma = \sqrt{N_\text{s}}/\|\mathbf{F}_\text{R}\mathbf{F}_\text{B}\|_F$ to force it to satisfy (8b). Moreover, when the number of RF chains is equal to the number of data streams ($L_\text{t} = N_\text{s}$) and the columns of the baseband precoding matrix are unitary to each other ($\mathbf{F}_\text{B}^H\mathbf{F}_\text{B} = \mathbf{I}$), $(\mathcal{P}_2)$ can be viewed as the orthonormal Procrustes problem (OPP) [10], [13] whose exact solution is $\mathbf{F}_\text{B} = \mathbf{V}_\text{F}\mathbf{U}_\text{F}^H$ resulting from the SVD of $\mathbf{F}_\text{opt}^H\mathbf{F}_\text{R} = \mathbf{U}_\text{F}\mathbf{\Sigma}_\text{F}\mathbf{V}_\text{F}^H$.

[1] $\text{vec}^{-1}(\cdot)$ means the inverse operation of matrix vectorization, namely, reshaping the $N_\text{t}L_\text{t} \times 1$ vector to a $N_\text{t} \times L_\text{t}$ matrix.

Particularly, with the assumption of semi-unitary of $\mathbf{F}_\text{R}$ [8], the power constraint can be simplified only respect to $\mathbf{F}_\text{B}$, i.e. $\|\mathbf{F}_\text{B}\|_F^2 = N_\text{s}$, which also facilitates the design of the hybrid precoders.

The SDR−AO algorithm (**Algorithm 1**) put forward in this subsection performs very well in dealing with $(\mathcal{P}_1)$, as we will show in section V-B. However, implementing SDR usually needs high computational complexity due to standard interior-point method, which will be analyzed in detail in Section V-A. According to this, we move now to explore other alternating optimization algorithms to design the hybrid precoders that require less complexity.

### C. ADMM−AO: THE ALTERNATING DIRECTION METHOD OF MULTIPLIERS BASED ALTERNATING OPTIMIZATION

We have seen in the previous section that, when RF precoder $\mathbf{F}_\text{R}$ is given, $(\mathcal{P}_3)$ can be reformulated as a UQP $(\mathcal{P}_5)$. As shown in [36], [37], the alternating direction method of multipliers (ADMM), which taking advantage of dual decomposition and augmented Lagrangian, is presented as an efficient algorithm to cope with UQP. For the sake of the usage of ADMM, we recast $(\mathcal{P}_5)$ as

$$(\mathcal{P}_7): \min_{\bar{\mathbf{x}},\bar{\mathbf{w}}} \frac{1}{2}\bar{\mathbf{x}}^H\mathbf{R}\bar{\mathbf{x}} + \tilde{I}_\mathcal{F}(\bar{\mathbf{w}}) \tag{18a}$$
$$\text{s.t. } \bar{\mathbf{x}} = \bar{\mathbf{w}}, \tag{18b}$$

where $\bar{\mathbf{x}}$ denotes the $n \times 1$ auxiliary variable, $\tilde{I}_\mathcal{F}(\cdot)$ stands for the indicator function [30] of the set $\mathcal{F}$ that is given by

$$\tilde{I}_\mathcal{F}(\bar{\mathbf{w}}) = \begin{cases} 0, & \bar{\mathbf{w}} \in \mathcal{F} \\ \infty, & \bar{\mathbf{w}} \notin \mathcal{F}. \end{cases} \tag{19}$$

As mentioned in the method of multipliers, the augmented Lagrangian function of $(\mathcal{P}_7)$ is formulated by

$$L_\mu(\bar{\mathbf{x}}, \bar{\mathbf{w}}, \mathbf{t}) = \frac{1}{2}\bar{\mathbf{x}}^H\mathbf{R}\bar{\mathbf{x}} + \tilde{I}_\mathcal{F}(\bar{\mathbf{w}}) + \mathbf{t}^H(\bar{\mathbf{x}} - \bar{\mathbf{w}}) + \frac{1}{2}\mu\|\bar{\mathbf{x}} - \bar{\mathbf{w}}\|_2^2, \tag{20}$$

where $\mathbf{t}$ is an $n \times 1$ vector of Lagrange Multipliers and $\mu > 0$ represents a scalar augmented Lagrangian parameter. In light of the ADMM approach [37], the solution of $(\mathcal{P}_7)$ is composed by the following iterations with respect to the augmented Lagrangian function (20), namely

$$\bar{\mathbf{x}}^{(k+1)} = \arg\min_{\bar{\mathbf{x}}} L_\mu\left(\bar{\mathbf{x}}, \bar{\mathbf{w}}^{(k)}, \mathbf{t}^{(k)}\right), \tag{21}$$
$$\bar{\mathbf{w}}^{(k+1)} = \arg\min_{\bar{\mathbf{w}}} L_\mu\left(\bar{\mathbf{x}}^{(k+1)}, \bar{\mathbf{w}}, \mathbf{t}^{(k)}\right), \tag{22}$$
$$\mathbf{t}^{(k+1)} = \mathbf{t}^{(k)} + \mu\left(\bar{\mathbf{x}}^{(k+1)} - \bar{\mathbf{w}}^{(k+1)}\right), \tag{23}$$

where $k$ is the iteration index. As we can see in (21)−(23), during each iteration, the augmented Lagrangian function is minimized alternately to harvest the optimization variable and the dual variable is updated with a step size $\mu$. Now,

we take (21) into account, which can be written in detail as

$$\min_{\bar{\mathbf{x}}} \frac{1}{2}\bar{\mathbf{x}}^H\mathbf{R}\bar{\mathbf{x}} + \left(\mathbf{t}^{(k)}\right)^H\left(\bar{\mathbf{x}} - \bar{\mathbf{w}}^{(k)}\right) + \frac{1}{2}\mu\left\|\bar{\mathbf{x}} - \bar{\mathbf{w}}^{(k)}\right\|_2^2. \tag{24}$$

Since the objective function $L_\mu\left(\bar{\mathbf{x}}, \bar{\mathbf{w}}^{(k)}, \mathbf{t}^{(k)}\right)$ in (24) is a convex quadratic function of $\bar{\mathbf{x}}$, we can find the minimum from the following optimality condition

$$\nabla_{\bar{\mathbf{x}}} L_\mu\left(\bar{\mathbf{x}}, \bar{\mathbf{w}}^{(k)}, \mathbf{t}^{(k)}\right) = (\mathbf{R} + \mu\mathbf{I}_n)\bar{\mathbf{x}} + \mathbf{t}^{(k)} - \mu\bar{\mathbf{w}}^{(k)} = 0, \tag{25}$$

where $\nabla_{\bar{\mathbf{x}}} L_\mu(\cdot)$ represents the gradient of function $L_\mu(\cdot)$. This equation yields the solution of (21), that is

$$\bar{\mathbf{x}}^{(k+1)} = (\mathbf{R} + \mu\mathbf{I}_n)^{-1}(\mu\bar{\mathbf{w}}^{(k)} - \mathbf{t}^{(k)}). \tag{26}$$

Subsequently, we will derive the solution of (22),which is restated by

$$\min_{\bar{\mathbf{w}}} \tilde{I}_{\mathcal{F}}(\bar{\mathbf{w}}) + \frac{1}{2}\mu\left\|\bar{\mathbf{w}} - \bar{\mathbf{x}}^{(k+1)} - \frac{1}{\mu}\mathbf{t}^{(k)}\right\|_2^2. \tag{27}$$

It is obvious that the solution of (22) is the projection of $\bar{\mathbf{x}}^{(k+1)} + \frac{1}{\mu}\mathbf{t}^{(k)}$ on $\mathcal{F}$. Thus, we have

$$\bar{\mathbf{w}}^{(k+1)} = \mathrm{P}_{\mathcal{F}}\left\{\bar{\mathbf{x}}^{(k+1)} + \frac{1}{\mu}\mathbf{t}^{(k)}\right\}, \tag{28}$$

where $\mathrm{P}_{\mathcal{F}}\{\mathbf{a}\}$ refers to the projection onto $\mathcal{F}$ for a arbitrary vector $\mathbf{a}$, while all non-zero elements in $\mathbf{a}$ are already normalized by itself.

So far, the design steps of ADMM approach to solve $(\mathcal{P}_5)$ has been finished, and we will reshape the RF precoding matrix $\mathbf{F}_{\mathrm{R}}$ by extracting the first $N_{\mathrm{t}}L_{\mathrm{t}}$ terms of $\bar{\mathbf{w}}$ and carrying out the matrix reconstruction operation $\mathbf{F}_{\mathrm{R}} = \mathrm{vec}^{-1}(\mathbf{w})$. Concerning about the baseband precoder design, we will adopt the simple scaling LS solution $\mathbf{F}_{\mathrm{B}} = \gamma\mathbf{F}_{\mathrm{R}}^{+}\mathbf{F}_{\mathrm{opt}}$ instead of the SDR method with high computational complexity shown in subsection III-B. From the above, the provided procedure for designing the hybrid precoders by virtue of ADMM−AO is summarized in **Algorithm 2**.

It is worth noting that choosing the penalty parameter $\mu$ appropriately is crucial to guarantee the convergence of ADMM in **Algorithm 2** and achieve satisfactory performance. Here, we choose the value $\mu = 10^2$ as a step size, which will also be exploited in the following simulations. Furthermore, we omit the analysis of convergence results of ADMM, cause that it has been discussed elaborately in [36], [37]. The ADMM−AO algorithm requires two-fold iterations, which may also lead to high complexity whenever a large number of iterations is appeared. Next, we will develop a more intuitive and straightforward alternating optimization algorithm to design the hybrid precoders.

### D. ACMF–AO: ANALYTICAL CONSTANT MODULUS FACTORIZATION BASED ALTERNATING OPTIMIZATION

Let us observe the problem $(\mathcal{P}_5)$ again. It is clear that the Hermitian matrix $\mathbf{R}$ in objective function is an nonnegative definite matrix ($\bar{\mathbf{w}}^H\mathbf{R}\bar{\mathbf{w}} \geq 0$) on account of the equivalence of $(\mathcal{P}_4)$ and $(\mathcal{P}_5)$. Furthermore, the rank of $n \times n$ matrix $\mathbf{R}$ satisfies the following inequality

$$\begin{aligned}\mathrm{rank}(\mathbf{R}) &= \mathrm{rank}([\mathbf{A}, -\mathbf{y}]) = \mathrm{rank}(\mathbf{A}) \\ &\leq N_{\mathrm{t}}N_{\mathrm{s}} < N_{\mathrm{t}}L_{\mathrm{t}} + 1 = n.\end{aligned} \tag{29}$$

Therefore, there must be a non-zero vector $\bar{\mathbf{w}}$ such that $\bar{\mathbf{w}}^H\mathbf{R}\bar{\mathbf{w}} = 0$, which means that the optimal value in $(\mathcal{P}_5)$ can be taken on the value zero. In particular, if we define the SVD of the singular matrix $\mathbf{R}$ as $\mathbf{R} = \mathbf{U}_{\mathrm{R}}\mathbf{\Sigma}_{\mathrm{R}}\mathbf{U}_{\mathrm{R}}^H$, the problem $(\mathcal{P}_5)$ can be converted to seek a vector that satisfies the following equation

$$\bar{\mathbf{w}} = \mathbf{U}_{\mathrm{R},2}\mathbf{p}, \tag{30a}$$

$$|\bar{\mathbf{w}}_i|^2 = 1, \quad i = 1, \ldots, n, \tag{30b}$$

where $\mathbf{U}_{\mathrm{R},2} \in \mathbb{C}^{n\times l}$ contains the last $l$ ($l = n - \mathrm{rank}(\mathbf{R})$) singular vectors of $\mathbf{R}$ corresponding to the singular value zero. $\mathbf{p}$ represents the $l \times 1$ complex coefficient vector. The problem (30a) of finding a linear combination of $l$ singular vectors to satisfy the unit-modulus constraint (30b) is described as a constant-modulus factorization problem [38], [39], which can be accurately solved by the analytic constant modulus factorization (ACMF) algorithm.

First, defining $\mathbf{U}_{\mathrm{R},2}^H = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$, the linear system (30a) can be decomposed into $\bar{\mathbf{w}}_i = \mathbf{u}_i^H\mathbf{p}, i = 1, \ldots, n$, where $\mathbf{u}_i^H$ is the $i^{\mathrm{th}}$ row of $\mathbf{U}_{\mathrm{R},2}$. After that, we will rewrite the constraint (30b) as

$$\begin{aligned}|\bar{\mathbf{w}}_i|^2 &= \mathbf{u}_i^H\mathbf{p}\mathbf{p}^H\mathbf{u}_i = \mathrm{vec}(\mathbf{u}_i^H\mathbf{p}\mathbf{p}^H\mathbf{u}_i) \\ &= (\mathbf{u}_i^T \otimes \mathbf{u}_i^H)\mathrm{vec}(\mathbf{p}\mathbf{p}^H)\end{aligned}$$

**Algorithm 2** ADMM−AO Based Hybrid Precoding Design

**Input**: $\mathbf{F}_{\mathrm{opt}}$

Initialize $\mathbf{F}_{\mathrm{R}}^{(0)}$ with random phases and set $\mathbf{t}^{(0)} = 0$, $\mu = 10^2$;
$\mathbf{F}_{\mathrm{B}}^{(0)} = (\mathbf{F}_{\mathrm{R}}^{(0)})^{+}\mathbf{F}_{\mathrm{opt}}$;
$\mathbf{F}_{\mathrm{B}}^{(0)} = \sqrt{N_{\mathrm{s}}}/\|\mathbf{F}_{\mathrm{R}}^{(0)}\mathbf{F}_{\mathrm{B}}^{(0)}\|_F \cdot \mathbf{F}_{\mathrm{B}}^{(0)}$;
$k = 0$; $i = 0$;
**repeat**
  Fix $\mathbf{F}_{\mathrm{B}}^{(k)}$, obtain $\bar{\mathbf{w}}$ by
  **repeat**
    $\bar{\mathbf{x}}^{(i+1)} = (\mathbf{R} + \mu\mathbf{I}_n)^{-1}(\mu\bar{\mathbf{w}}^{(i)} - \mathbf{t}^{(i)})$;
    $\bar{\mathbf{w}}^{(i+1)} = P_{\mathcal{F}}\{\bar{\mathbf{x}}^{(i+1)} - \frac{1}{\mu}\mathbf{t}^{(i)}\}$;
    $\mathbf{t}^{(i+1)} = \mathbf{t}^{(i)} + \mu\left(\bar{\mathbf{x}}^{(i+1)} - \bar{\mathbf{w}}^{(i+1)}\right)$;
    $i = i + 1$;
  **until** a stopping criterion triggers
  Update $\mathbf{F}_{\mathrm{R}}^{(k+1)}$ by matrix reconstruction;
  Fix $\mathbf{F}_{\mathrm{R}}^{(k+1)}$, update $\mathbf{F}_{\mathrm{B}}^{(k+1)}$ with $\mathbf{F}_{\mathrm{B}}^{(k+1)} = (\mathbf{F}_{\mathrm{R}}^{(k+1)})^{+}\mathbf{F}_{\mathrm{opt}}$;
  $\mathbf{F}_{\mathrm{B}}^{(k+1)} = \sqrt{N_{\mathrm{s}}}/\|\mathbf{F}_{\mathrm{R}}^{(k+1)}\mathbf{F}_{\mathrm{B}}^{(k+1)}\|_F \cdot \mathbf{F}_{\mathrm{B}}^{(k+1)}$;
  $k = k + 1$;
**until** a stopping criterion triggers

**Output**: $\mathbf{F}_{\mathrm{R}}, \mathbf{F}_{\mathrm{B}}$

$$= (\mathbf{u}_i^T \otimes \mathbf{u}_i^H)(\mathbf{p}^* \otimes \mathbf{p}) = 1. \tag{31}$$

Stacking all $n$ such constraints into a single linear system results in

$$\mathbf{T}\mathbf{v} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \tag{32}$$

where

$$\mathbf{T} = \begin{bmatrix} \mathbf{u}_1^T \otimes \mathbf{u}_1^H \\ \vdots \\ \mathbf{u}_n^T \otimes \mathbf{u}_n^H \end{bmatrix}, \quad \mathbf{v} = \mathbf{p}^* \otimes \mathbf{p}. \tag{33}$$

Choosing a simple Householder transformation matrix or normalized discrete Fourier transform (DFT) matrix as an $n \times n$ unitary matrix $\mathbf{Q}$, such that

$$\mathbf{Q}\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} \sqrt{n} \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{34}$$

Accordingly, the linear system (32) will be precisely equivalent to the following equations

$$\hat{\mathbf{t}}_1^H \mathbf{v} = \sqrt{n}, \tag{35a}$$
$$\hat{\mathbf{T}}\mathbf{v} = \mathbf{0}, \tag{35b}$$

where the $l^2 \times 1$ vector $\hat{\mathbf{t}}_1$ and $(n-1) \times l^2$ matrix $\hat{\mathbf{T}}$ are determined by $\mathbf{Q}\mathbf{T} = [\hat{\mathbf{t}}_1, \hat{\mathbf{T}}^H]^H$. Notice that the linear equation (35a) is equivalent to

$$\|\mathbf{p}\|_2^2 = n. \tag{36}$$

The proof of the equivalence between (35a) and (36) can be found in Appendix A. Hence, the equation (35a) can always be satisfied by scaling the solution of the homogeneous linear equation (35b). Now, the nonhomogeneous linear system (32) with certain structure will be restated as

$$\begin{cases} \mathbf{T}\mathbf{v} \hat{=} 0 \\ \mathbf{v} = \mathbf{p}^* \otimes \mathbf{p} \end{cases} \tag{37}$$

The exact solution of (37) can be obtained by solving a simultaneous diagonalization problem, when $(n > l^2)$ is supposed (see [38] for more details). Here, we adopt the following procedures to obtain an approximation:

1) Perform the SVD of $\hat{\mathbf{T}}$ ($\hat{\mathbf{T}} = \mathbf{U}_\mathrm{T}\mathbf{\Sigma}_\mathrm{T}\mathbf{V}_\mathrm{T}^H$) and select $\mathbf{v}_\mathrm{T}$ as the right singular vector in $\mathbf{V}_\mathrm{T}$ associated with the singular value zero.
2) Reshape the $l^2 \times 1$ vector $\mathbf{v}_\mathrm{T}$ into a $l \times l$ matrix, $\mathbf{\Gamma} = \mathrm{vec}^{-1}(\mathbf{v}_\mathrm{T})$. The coefficient vector $\mathbf{p}$ equal to the singular vector of $\mathbf{\Gamma} + \mathbf{\Gamma}^H$ corresponding to the largest singular value.
3) Scale $\mathbf{p}$ by $\sqrt{n}/\|\mathbf{p}\|_2$ and set $\mathbf{w} = \hat{\mathbf{U}}_{\mathrm{R},2}\mathbf{p}$. Then, the vector $\bar{\mathbf{w}}$ is obtained by setting the magnitude of all elements of $\hat{\mathbf{w}}$ equal to unity.

Hence, by using the ACMF algorithm, the problem $(\mathcal{P}_5)$ can be solved again. In next, the reconstruction of the RF precoding matrix $\mathbf{F}_\mathrm{R}$ and design of the baseband precoding matrix $\mathbf{F}_\mathrm{B}$ is similar to the steps elaborated in subsection III-B and III-C. The detailed algorithm design procedure of ACMF−AO based hybrid precoders are omitted, since most of the processes are the same as **Algorithm 2** except for the solution of $\bar{\mathbf{w}}$.

Until now, we have put forward three alternating optimization methods based on different principles to design hybrid precoding matrix in a narrowband channel. In the following discussion, we will analyze the design of Hybrid Precoder in Wideband MmWave Systems.

## IV. HYBRID PRECODING IN WIDEBAND MmWave SYSTEMS

In this section, the proposed AO algorithms are extended to mmWave wideband systems. As shown in [10], the hybrid precoder design problem in OFDM-based wideband mmWave MIMO systems can be written as

$$\min_{\mathbf{F}_\mathrm{R}, \mathbf{F}_\mathrm{B}[k]} \sum_{k=1}^{K} \left\| \mathbf{F}_\mathrm{opt}[k] - \mathbf{F}_\mathrm{R}\mathbf{F}_\mathrm{B}[k] \right\|_F^2 \tag{38a}$$
$$\text{s.t.} \quad \mathbf{F}_\mathrm{R}(i,j) \in \mathcal{F}, \quad \forall i,j, \tag{38b}$$
$$\|\mathbf{F}_\mathrm{R}\mathbf{F}_\mathrm{B}[k]\|_F^2 = N_\mathrm{s}, \tag{38c}$$

where $k \in [1, K]$ is the subcarrier index, $\mathbf{F}_\mathrm{opt}[k]$ and $\mathbf{F}_\mathrm{B}[k]$ represent the optimal unconstrained digital precoder and baseband precoder for the $k^{\text{th}}$ subcarrier, respectively. It is worth noting that the analog precoder is constant over the subcarriers, since it is implemented after inverse fast Fourier transform (IFFT) processing. Similar to the narrowband case, the alternating optimization algorithm is still one of the good choices to solve the problem (38). Particularly, the digital precoder for each subcarrier should be updated in parallel. Hence, the digital precoding matrix design problem for the $k^{\text{th}}$ subcarrier is the same as $(\mathcal{P}_2)$. Here, we adopt the scaling LS solution $\mathbf{F}_\mathrm{B}[k] = \gamma_k \mathbf{F}_\mathrm{R}^+ \mathbf{F}_\mathrm{opt}[k]$ ($\gamma_k = \sqrt{N_\mathrm{s}}/\|\mathbf{F}_\mathrm{R}\mathbf{F}_\mathrm{B}[k]\|_F$). After that, the analog precoder design problem will be accordingly formulated as

$$\min_{\mathbf{F}_\mathrm{R}} \sum_{k=1}^{K} \left\| \mathbf{F}_\mathrm{opt}[k] - \mathbf{F}_\mathrm{R}\mathbf{F}_\mathrm{B}[k] \right\|_F^2 \tag{39a}$$
$$\text{s.t.} \quad |\mathbf{F}_\mathrm{R}(i,j)|^2 = 1, \quad \forall i,j. \tag{39b}$$

The problem is similar to $(\mathcal{P}_3)$ except for the summation in the objective function. Next, we will execute matrix vectorization to transform it into an UQP problem. Denoting $\mathbf{y}[k] = \mathrm{vec}\left(\mathbf{F}_\mathrm{opt}[k]\right)$, $\mathbf{A}[k] = \mathbf{F}_\mathrm{B}^T[k] \otimes \mathbf{I}_{N_\mathrm{t}}$, and $\mathbf{w} = \mathrm{vec}\left(\mathbf{F}_\mathrm{R}\right)$, then the objective function in (39a) is transformed into $\sum_{k=1}^{K} \|\mathbf{y}[k] - \mathbf{A}[k]\mathbf{w}\|_2^2$. With the introduction of auxiliary variable $t$, the problem (39) is finally homogenized as

$$\min_{\bar{\mathbf{w}}} \ \bar{\mathbf{w}}^H \sum_{k=1}^{K} \mathbf{R}[k]\bar{\mathbf{w}} \tag{40a}$$
$$\text{s.t.} \quad |\bar{\mathbf{w}}_i|^2 = 1, \quad i = 1, \ldots, N_\mathrm{t}L_\mathrm{t} + 1, \tag{40b}$$

**TABLE 1. Complexity comparison of the proposed and competing algorithms.**

| Algorithm | Computational Complexity |
|---|---|
| SDR−AO | $\mathcal{O}\big(k(L_{\rm t}^7(N_{\rm t}^7+N_{\rm s}^7)+L_{\rm t}^3(N_{\rm t}^3+N_{\rm s}^3))\big)$ |
| ADMM−AO | $\mathcal{O}\big(k_o k_i(N_{\rm t}^3L_{\rm t}^3+N_{\rm t}^2L_{\rm t}^2)\big)$ |
| ACMF−AO | $\mathcal{O}\big(k(N_{\rm t}^3L_{\rm t}^3+(l^2+l)N_{\rm t}^2L_{\rm t}^2)\big)$ |
| OMP [9] | $\mathcal{O}\big(N_{\rm t}^2L_{\rm t}N_{\rm s}\big)$ |
| PE−AltMin [10] | $\mathcal{O}\big(kN_{\rm t}L_{\rm t}^2\big)$ |
| ICDA [11] | $\mathcal{O}\big(N_{\rm t}^2L_{\rm t}^2+N_{\rm t}L_{\rm t}^3\big)$ |
| HD−LSR [13] | $\mathcal{O}\big(k(N_{\rm t}L_{\rm t}^2N_{\rm s}+N_{\rm t}L_{\rm t}N_{\rm s})\big)$ |
| GP−AltMin [15] | $\mathcal{O}\big(k(N_{\rm t}^3L_{\rm t}^3+N_{\rm t}^3L_{\rm t}^2N_{\rm s})\big)$ |

where

$$\mathbf{R}[k]=\begin{bmatrix}\mathbf{A}[k]^H\mathbf{A}[k] & -\mathbf{A}[k]^H\mathbf{y}[k]\\ -\mathbf{y}[k]^H\mathbf{A}[k] & \mathbf{y}[k]^H\mathbf{y}[k]\end{bmatrix},\ \bar{\mathbf{w}}=\begin{bmatrix}\mathbf{w}\\ t\end{bmatrix}. \quad (41)$$

The above discussion demonstrates that the three aforementioned AO algorithms are also applicable to the wideband mmWave systems. In the next section, we will analyze the performance of different hybrid precoding algorithms from the perspective of computational complexity and spectral efficiency.

## V. RESULTS

### A. COMPUTATIONAL COMPLEXITY ANALYSES

In this subsection, we discuss the computational complexity with respect to the aforementioned SDR−AO, ADMM−AO and ACMF−AO hybrid precoding algorithms and compare with other novel algorithms. Specifically, for the SDR−AO algorithm, the computational complexity is dominated by convex optimization toolbox with a complexity of $\mathcal{O}(n^{3.5})$ as shown in [33], in which the interior-point method is utilized. However, the SDR processing enhances the dimension of problem to $n^2$, which means that $\mathcal{O}(n^7)$ floating point operations is required to solve $(\mathcal{P}_6)$. Similarly, solving (16) to obtain $\mathbf{W}$ requires $\mathcal{O}(m^7)$ computations. In addition to this, during the rank-1 approximation for updating the matrix $\mathbf{F}_{\rm R}$, $\mathbf{F}_{\rm B}$, the SVD should be executed with complexity $\mathcal{O}(n^3)$, $\mathcal{O}(m^3)$. Therefore, the overall computational complexity of the SDR−AO algorithm presented in **Algorithm 1** is given by $\mathcal{O}\big(k(L_{\rm t}^7(N_{\rm t}^7+N_{\rm s}^7)+L_{\rm t}^3(N_{\rm t}^3+N_{\rm s}^3))\big)$, where $k$ is the iteration index. Here and next, "1" in $n$, $m$ ($n=N_{\rm t}L_{\rm t}+1$, $m=L_{\rm t}N_{\rm s}+1$) is discarded so that we can compare the complexity of different algorithms directly.

For the ADMM−AO based hybrid precoding design, the calculation of $\bar{\mathbf{w}}$ includes matrix-vector multiplication and matrix inversion operations which require complexity of $\mathcal{O}(n^2)$ and $\mathcal{O}(n^3)$, respectively. Moreover, updating the matrix $\mathbf{F}_{\rm B}$ with the scaling LS solution involves the Moore-Penrose inversion (here, the Moore-Penrose inversion of $\mathbf{F}_{\rm R}$ is implemented by $(\mathbf{F}_{\rm R}^H\mathbf{F}_{\rm R})^{-1}\mathbf{F}_{\rm R}^H$ virtually) and matrix-matrix multiplication operations, which bring about $\mathcal{O}\big(L_{\rm t}^2N_{\rm t}\big)$ and $\mathcal{O}\big(L_{\rm t}N_{\rm t}N_{\rm s}\big)$ computations. However, since $L_{\rm t}N_{\rm t}N_{\rm s}\le L_{\rm t}^2N_{\rm t}\ll n^2$, we will leave this calculation out in the final complexity result. In short, the computational complexity of the ADMM−AO algorithm shown in **Algorithm 2** is $\mathcal{O}\big(k_ok_i(N_{\rm t}^3L_{\rm t}^3+N_{\rm t}^2L_{\rm t}^2)\big)$, where $k_o$ and $k_i$ are the numbers of external and internal iterations.

For the ACMF−AO hybrid precoding algorithm, the complexity is dominated by the SVD and matrix-matrix multiplication operations. Implementing the SVD on matrices $\mathbf{R}$ and $\hat{\mathbf{T}}$ requires $\mathcal{O}(n^3)$ and $\mathcal{O}(l^2n^2)$ computations. The other multiplication operation parts have complexity $\mathcal{O}(ln^2)$. In consequence, the ACMF−AO algorithm requires computational complexity of $\mathcal{O}\big(k(N_{\rm t}^3L_{\rm t}^3+(l^2+l)N_{\rm t}^2L_{\rm t}^2)\big)$ totally.

At last, the computational complexity of the proposed three algorithms and the novel algorithms from the previous literature are listed in Table 1. Compared with the OMP [9], PE−AltMin [10], ICDA [11] and HD−LSR [13] algorithms, the complexity of presented ADMM−AO and ACMF−AO algorithms are one order of magnitude higher, especially the SDR−AO algorithm show extremely high computational complexity. However, the complexity of ADMM−AO, ACMF−AO is comparable to GP−AltMin [15], while ADMM−AO may slightly higher due to the two-fold iterations.

### B. SIMULATION RESULTS

In this section, we evaluate the spectral efficiency performance of the presented hybrid precoding algorithm by numerical simulations, where the OMP [9], PE−AltMin [10], ICDA [11], HD−LSR [13] and GP−AltMin [15] algorithms are considered as the competitors and the optimal digital precoding scheme is employed as the benchmark. The propagation environment is modeled as $N_{\rm c}=8$ cluster and $N_{\rm p}=10$ path with Laplacian distributed with an angle spread of 5°. The AoA/AoD azimuths of the cluster are assumed to be uniformly distributed in $[0,2\pi]$ and channel path gains are assumed to be Gaussian distribution with the variance $\sigma_{\alpha,i}^2=1$. The total power constraint is fixed and power allocation for each stream is equal and signal-to-noise ratio (SNR) is defined as SNR $=P/\sigma_{\rm n}^2$. Finally, all achievable simulation results are averaged over 100 random channel realizations.

Fig. 1, 2 illustrate the spectral efficiency for the proposed algorithms and other state-of-the-art algorithms in a $64\times16$ mmWave MIMO system with $L_{\rm t}=N_{\rm s}=4(8)$. Obviously, over the whole SNR range in consideration, the best performing algorithm is the proposed SDR−AO algorithm, followed closely by PE−AltMin, ACMF−AO, and GP−AltMin, respectively, and there is only a small gap of them. In the case of small data streams and RF chains, the proposed ADMM−AO algorithm achieves better performance than the OMP algorithm in the low SNR range ($-20$ − 0 dB) and less performance in the medium SNR range (0 − 10 dB). Note that mmWave MIMO systems usually operate in low and medium SNR regimes. However, as shown in Fig. 2, when the amount of data streams and RF chains increases properly, the performance of ADMM−AO

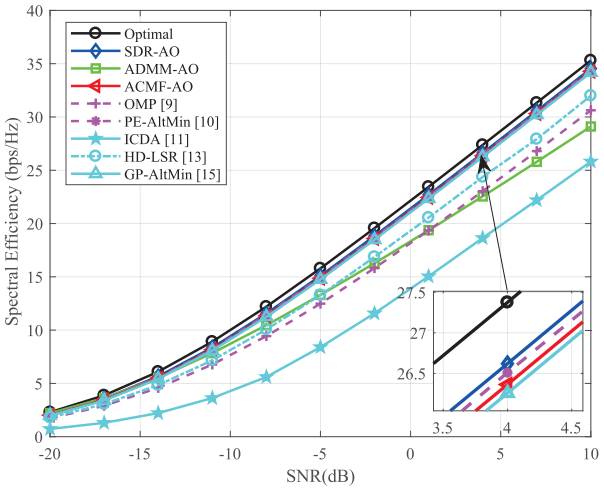



**FIGURE 1.** Spectral efficiency versus SNR for different precoding algorithms in a 64×16 mmWave MIMO system with $L_t = N_s = 4$.

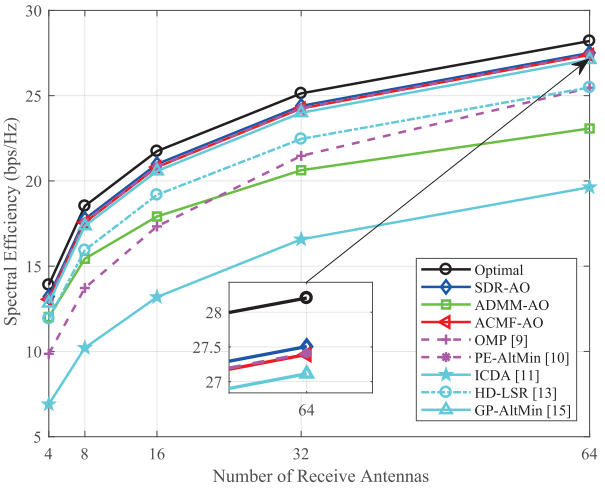



**FIGURE 3.** Spectral efficiency versus number of receive antennas for different precoding algorithms ($N_t = 64$, $L_t = N_s = 4$, SNR = 0 dB).

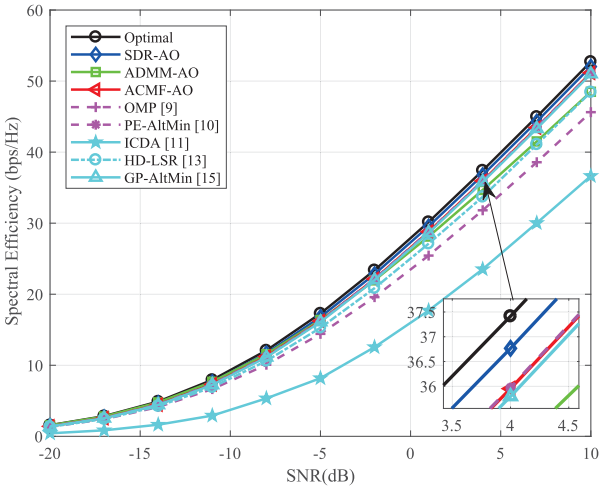



**FIGURE 2.** Spectral efficiency versus SNR for different precoding algorithms in a 64 × 16 mmWave MIMO system with $L_t = N_s = 8$.

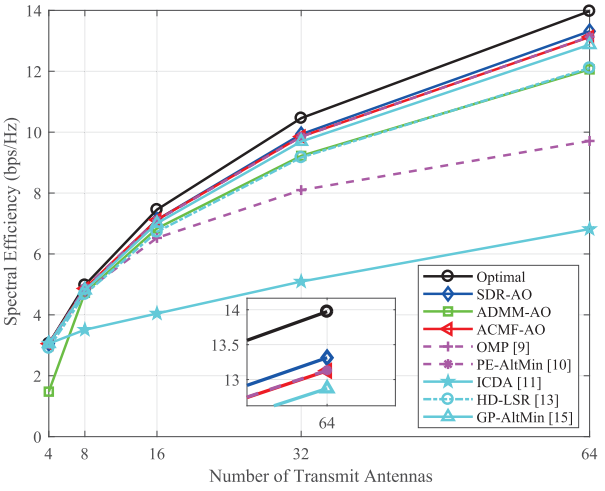



**FIGURE 4.** Spectral efficiency versus number of transmit antennas for different precoding algorithms ($N_r = 4$, $L_t = N_s = 4$, SNR = 0 dB).

algorithm is close to that of GP−AltMin algorithm and better than OMP algorithm. In addition, HD−LSR algorithm is always superior to OMP algorithm and inferior to GP−AltMin algorithm and the spectral efficiency obtained by the ICDA algorithm is far smaller than the other algorithms.

In order to examine the impacts of the transceiver antennas, we demonstrate the spectral efficiency versus the number of receive antennas $N_r$ and transmit antennas $N_t$ with $L_t = N_s = 4$, SNR $= 0$ dB in Fig. 3 and 4, respectively. It is observed that the SDR−AO algorithm consistently remains the closest to the optimal digital precoding scheme as receive/transmit antenna increases, still followed closely by PE−AltMin, ACMF−AO, and GP−AltMin, that implies the versatility of them in mmWave MIMO systems. Nevertheless, when the number of receive antennas is much smaller than that of transmit antennas, the proposed ADMM−AO algorithm is distinctly superior to the OMP algorithm and can almost achieve the same spectral efficiency as HD−LSR, as shown in Fig. 4, which indicates that the proposed ADMM−AO algorithm is applicable to the case $N_t \gg N_r$. Furthermore, the ICDA algorithm is sensitive to transceiver antennas since it shows a large performance gap with other algorithms.

The above simulation results show the superiority of the proposed algorithms when equal number of RF chains and data streams are employed. Next, we will depict the spectral efficiency of the above algorithms vesus the number of data streams $N_s$ in Fig. 5, where $L_t = 8$ and SNR $= 0$ dB. Observe that PE-AltMin is not applied because the orthogonal structure of digital precoder imposes same number of RF chains and data streams, while the constraint $n > l^2$ also makes $L_t = N_s$ the best choice for applying ACMF−AO, therefore, ACMF−AO is also not involved. As shown in Fig. 5, the HD−LSR algorithm and proposed ADMM−AO algorithm outperform the SDR−AO algorithm and other

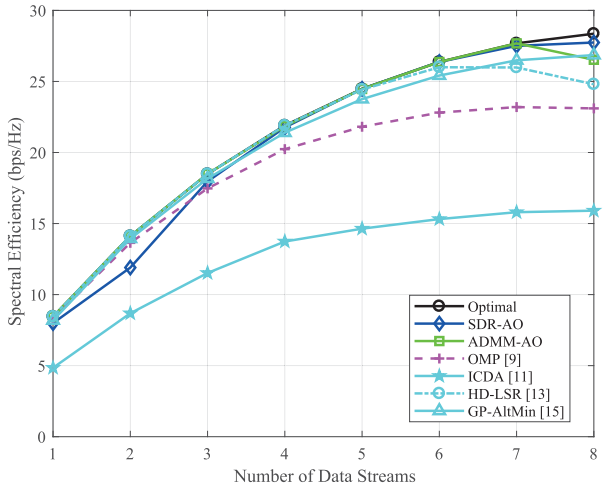



**FIGURE 5.** Spectral efficiency versus number of data streams for different precoding algorithms ($N_t \times N_r = 64 \times 16$, $L_t = 8$, SNR = 0 dB).

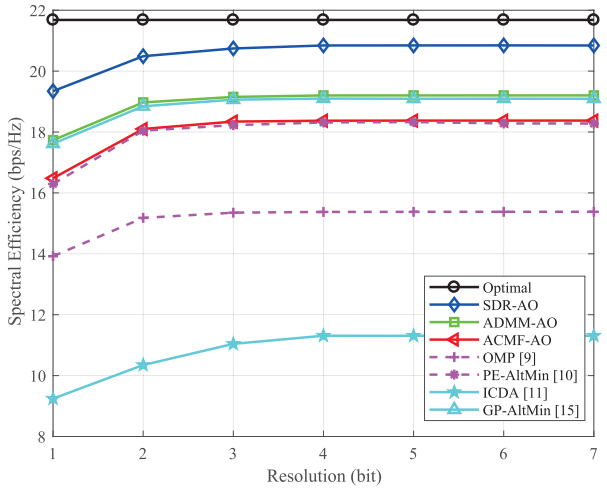



**FIGURE 6.** Spectral efficiency versus resolution of PSs for different precoding algorithms ($N_t \times N_r = 64 \times 16$, $L_t = N_s = 4$, SNR = 0 dB).

algorithms when the number of data streams is much less than RF chains. However, when the number of data streams is equal to the number of RF chains, the performance of ADMM−AO and HD−LSR significantly deteriorates, which implies that these two algorithms are more suitable for low data streams transmission scenarios. Finally, we examine the effects of the resolution of PSs in Fig. 6, when designing the analog component of the hybrid precoder and combiner. As expected, the spectral efficiency will improve as the resolution of PS increases. It can be observed that the proposed three algorithm performs better than the PE−AltMin algorithm for all considered PS resolutions. For the presented algorithms, employing only $B = 4$ bits is sufficient to closely approach the performance of the optimal full-digital case.

In summary, we can draw the following conclusions. Our proposed three algorithms apparently outperform the OMP and ICDA algorithm, in any case. Regarding spectral efficiency performance, SDR−AO performs best but at cost of extremely high computational complexity. When the number of transmit antennas is much larger than that of receive antennas (which is always the case in massive MIMO) or the amount of data streams is small, ADMM−AO is always preferred, since it performs very close to SDR−AO with a relatively lower complexity. Furthermore, ACMF−AO is a better choice considering the trade-off between computational complexity and spectral efficiency performance, when equal number of RF chains and data streams are employed.

## VI. CONCLUSIONS

In this paper, we put forward several alternating optimization strategies to design hybrid precoding for P2P mmWave MIMO systems. Firstly, the original matrix factorization problem is iteratively decoupled into a nonconvex QCQP and an ULS problem after matrix vectorization. The non-convex QCQP can be transformed into a SDP and then solved by the SDR method or be solved by scaling the unconstrained LS problem. Afterwards, we focus on developing three different algorithms to solve the ULS problem. Finally, the computational complexity analyses and simulation results demonstrate that, in each case, the proposed algorithms are superior to previously proposed algorithms from the literature, especially the SDR−AO can achieve near-optimal spectral efficiency but at cost of extremely high complexity. The ADMM−AO is preferred in the case that when the number of transmit antennas is much larger than that of receive antennas or the amount of data streams is small. When equal number of RF chains and data streams are employed, ACMF−AO is a better choice. Furthermore, the proposed methods are also applicable to the design of hybrid combiner at receiver due to the similar structure and can be well extended to wideband mmWave systems.

## APPENDIX A
## PROOF OF THE EQUIVALENCE BETWEEN (35a) AND (36)

Defining $\mathbf{u}_i^H = [u_{i1}, \ldots, u_{il}]$, then $\mathbf{T}$ can be expressed as

$$\mathbf{T} = \begin{bmatrix} u_{11}^* u_{11} & \cdots & u_{11}^* u_{1l} & \cdots & u_{1l}^* u_{11} & \cdots & u_{1l}^* u_{1l} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ u_{n1}^* u_{n1} & \cdots & u_{n1}^* u_{nl} & \cdots & u_{nl}^* u_{n1} & \cdots & u_{nl}^* u_{nl} \end{bmatrix}. \tag{42}$$

Denoting the first row of the unitary matrix $\mathbf{Q}$ as $\tilde{\mathbf{q}}_1^H$, which much be equal to $\frac{1}{\sqrt{n}}[1, \ldots, 1]$, thus, $\hat{\mathbf{t}}_1^H$ can be expressed as

$$\hat{\mathbf{t}}_1^H = \tilde{\mathbf{q}}_1^H \mathbf{T} = \frac{1}{\sqrt{n}} \left[ \tilde{\mathbf{u}}_1^H \tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_1^H \tilde{\mathbf{u}}_l, \ldots, \tilde{\mathbf{u}}_l^H \tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_l^H \tilde{\mathbf{u}}_l \right], \tag{43}$$

where $\tilde{\mathbf{u}}_j = [u_{1j}, \ldots, u_{nj}]^T$ is the $j^{\text{th}}$ column of $\mathbf{U}_{\text{R},2}$. Employing the semi-unitary property of $\mathbf{U}_{\text{R},2}$ ($\mathbf{U}_{\text{R},2}^H \mathbf{U}_{\text{R},2} = \mathbf{I}_l$) and reshaping $\hat{\mathbf{t}}_1$ as a matrix gives

$$\tilde{\mathbf{T}}_t = \text{vec}^{-1}(\hat{\mathbf{t}}_1) = \frac{1}{\sqrt{n}} \mathbf{I}_l. \tag{44}$$

Accordingly, we have

$$\hat{\mathbf{t}}_1^H \mathbf{v} = (\mathbf{p}^* \otimes \mathbf{p})^T \mathrm{vec}(\tilde{\mathbf{T}}_t) = \mathbf{p}^T \tilde{\mathbf{T}}_t \mathbf{p}^* = \frac{1}{\sqrt{n}} \mathbf{p}^T \mathbf{p}^* = \frac{1}{\sqrt{n}} \|\mathbf{p}\|_2^2 = \sqrt{n}. \quad (45)$$

Finally, we arrive at the desired result in (36).

## REFERENCES

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

[2] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: The next wireless revolution?" *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 56–62, Sep. 2014.

[3] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[4] R. W. Heath, N. Gonzalez-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.

[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] L. Fan, S. Jin, C.-K. Wen, and H. Zhang, "Uplink achievable rate for massive MIMO systems with low-resolution ADC," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2186–2189, Dec. 2015.

[7] Q. Hou, R. Wang, E. Liu, and D. Yan, "Hybrid precoding design for MIMO system with one-bit ADC receivers," *IEEE Access*, vol. 6, pp. 48478–48488, 2018.

[8] J. Mo, A. Alkhateeb, S. Abu-Surra, and R. W. Heath, Jr., "Hybrid architectures with few-bit ADC receivers: Achievable rates and energy-rate tradeoffs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2274–2287, Apr. 2017.

[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] 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.

[11] 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.

[12] W. Ni, X. Dong, and W.-S. Lu, "Near-optimal hybrid processing for massive MIMO systems via matrix decomposition," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 3922–3933, Aug. 2017.

[13] C. Rusu, R. Mendez-Rial, N. Gonzalez-Prelcic, and R. W. Heath, Jr., "Low complexity hybrid precoding strategies for millimeter wave communication systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8380–8393, Dec. 2016.

[14] J. Jin, Y. R. Zheng, W. Chen, and C. Xiao, "Hybrid precoding for millimeter wave MIMO systems: A matrix factorization approach," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 3327–3339, May 2018.

[15] J.-C. Chen, "Gradient projection-based alternating minimization algorithm for designing hybrid beamforming in millimeter-wave MIMO systems," *IEEE Commun. Lett.*, vol. 23, no. 1, pp. 112–115, Jan. 2019.

[16] 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.

[17] 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.

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

[19] R. A. Stirling-Gallacher and M. S. Rahman, "Linear MU-MIMO precoding algorithms for a millimeter wave communication system using hybrid beam-forming," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2014, pp. 5449–5454.

[20] 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.

[21] J. Du, W. Xu, H. Shen, X. Dong, and C. Zhao, "Hybrid precoding architecture for massive multiuser MIMO with dissipation: Sub-connected or fully connected structures?" *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5465–5479, Aug. 2018.

[22] X. Gao, L. Dai, S. Han, C.-L. I, and R. W. Heath, Jr., "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.

[23] R. Magueta, D. Castanheira, A. Silva, R. Dinis, and A. Gameiro, "Hybrid multi-user equalizer for massive MIMO millimeter-wave dynamic subconnected architecture," *IEEE Access*, vol. 7, pp. 79017–79029, Jun. 2019.

[24] S. Park, A. Alkhateeb, and R. W. Heath, Jr., "Dynamic subarrays for hybrid precoding in wideband mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2907–2920, May 2017.

[25] D. Castanheira, P. Lopes, A. Silva, and A. Gameiro, "Hybrid beamforming designs for massive MIMO millimeter-wave heterogeneous systems," *IEEE Access*, vol. 5, pp. 21806–21817, 2017.

[26] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing—The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[27] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda, "Hybrid beamforming for massive MIMO: A survey," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, Sep. 2017.

[28] R. Mendez-Rial, C. Rusu, N. Gonzalez-Prelcic, A. Alkhateeb, and R. W. Heath, Jr., "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, Jan. 2016.

[29] Z. Wang, M. Li, Q. Liu, and A. L. Swindlehurst, "Hybrid precoder and combiner design with low-resolution phase shifters in mmWave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 256–269, May 2018.

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

[31] H. Lütkepohl, *Handbook of Matrices*. Hoboken, NJ, USA: Wiley, 1996.

[32] J. Tranter, N. D. Sidiropoulos, X. Fu, and A. Swami, "Fast unit-modulus least squares with applications in beamforming," *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2875–2887, Jun. 2017.

[33] Z.-Q. Luo, W.-K. Ma, A. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.

[34] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

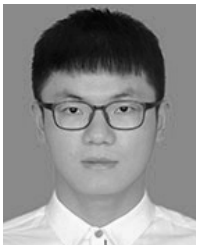
[35] W. Ai, Y. Huang, and S. Zhang, "New results on hermitian matrix rank-one decomposition," *Math. Program.*, vol. 128, nos. 1–2, pp. 253–283, Jun. 2011.

[36] C. G. Tsinos and B. Ottersten, "An efficient algorithm for unit-modulus quadratic programs with application in beamforming for wireless sensor networks," *IEEE Signal Process. Lett.*, vol. 25, no. 2, pp. 169–173, Feb. 2018.

[37] S. Boyd, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[38] A.-J. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Trans. Signal Process.*, vol. 44, no. 5, pp. 1136–1155, May 1996.

[39] F. Jiang, J. Chen, and A. L. Swindlehurst, "Estimation in phase-shift and forward wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 61, no. 15, pp. 3840–3851, Aug. 2013.



**XU QIAO** received the B.S. degree from the Nanjing Institute of Technology, Nanjing, China, in 2017. He is currently pursuing the M.S. degree with the College of Communication and Information Engineering, Nanjing University of Posts and Telecommunications (NJUPT), Nanjing. His research interests include wireless communications, massive MIMO, and mmWave hybrid precoding.

**YAO ZHANG** (Student Member, IEEE) received the B.S. degree from the College of Computer Science and Technology, Qingdao University, China, in 2016. He is currently pursuing the Ph.D. degree with the Department of Communication and Information Engineering, Nanjing University of Posts and Telecommunications. His current research interests include massive (large-scale) MIMO systems and performance analysis of fading channels.

**MENG ZHOU** (Student Member, IEEE) received the M.S. degree from Northwest Normal University, Lanzhou, China, in 2018. He is currently pursuing the Ph.D. degree with the College of Communication and Information Engineering, Nanjing University of Posts and Telecommunications (NJUPT), Nanjing, China. His research interest includes wireless communications, with a special interest of massive MIMO, physical layer security, and HetNets in 5G and beyond.

**LONGXIANG YANG** is currently with the College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications (NJUPT), Nanjing, China. He is also a Full Professor and a Doctoral Supervisor with NJUPT, where he is also the Vice Head of the College of Telecommunications and Information Engineering. He has fulfilled multiple National Natural Science Foundation projects of China. He has authored or coauthored over 100 technical papers published in various journals and conferences. His research interests include cooperative communication, network coding, wireless communication theory, and the Internet of Things.