Alternating Minimization Algorithms for Hybrid Precoding in Millimeter Wave MIMO Systems

Xianghao Yu, *Student Member, IEEE*, Juei-Chin Shen, *Member, IEEE*, Jun Zhang, *Senior Member, IEEE*,
and Khaled B. Letaief, *Fellow, IEEE*

Abstract—Millimeter wave (mmWave) communications has been regarded as a key enabling technology for 5G networks, as it offers orders of magnitude greater spectrum than current cellular bands. In contrast to conventional multiple-input–multiple-output (MIMO) systems, precoding in mmWave MIMO cannot be performed entirely at baseband using digital precoders, as only a limited number of signal mixers and analog-to-digital converters can be supported considering their cost and power consumption. As a cost-effective alternative, a hybrid precoding transceiver architecture, combining a digital precoder and an analog precoder, has recently received considerable attention. However, the optimal design of such hybrid precoders has not been fully understood. In this paper, treating the hybrid precoder design as a matrix factorization problem, effective alternating minimization (AltMin) algorithms will be proposed for two different hybrid precoding structures, i.e., the fully-connected and partially-connected structures. In particular, for the fully-connected structure, an AltMin algorithm based on manifold optimization is proposed to approach the performance of the fully digital precoder, which, however, has a high complexity. Thus, a low-complexity AltMin algorithm is then proposed, by enforcing an orthogonal constraint on the digital precoder. Furthermore, for the partially-connected structure, an AltMin algorithm is also developed with the help of semidefinite relaxation. For practical implementation, the proposed AltMin algorithms are further extended to the broadband setting with orthogonal frequency division multiplexing modulation. Simulation results will demonstrate significant performance gains of the proposed AltMin algorithms over existing hybrid precoding algorithms. Moreover, based on the proposed algorithms, simulation comparisons between the two hybrid precoding structures will provide valuable design insights.

Index Terms—Alternating minimization, hybrid precoding, low-complexity, manifold optimization, millimeter wave communications, semidefinite relaxation.

I. Introduction

THE capacity of wireless networks has to exponentially increase to meet the explosive demands for high-data-rate multimedia access. In particular, the upcoming 5G networks aim at carrying out the projected 1000X increase in capacity by 2020 [2]. One way to boost the capacity is to improve the spectral efficiency through physical layer techniques, such as massive multiple-input-multiple-output (MIMO) and advanced channel coding [3]. Further improvement in area spectral efficiency can be achieved by network densification, such as deploying small cells [4], [5] and allowing device-to-device (D2D) communications [6], and enabling advanced cooperation, such as Cloud-RANs [7], [8]. Nevertheless, the spectrum crunch in current cellular systems brings a fundamental bottleneck for the further capacity increase. Thus, it is critical to exploit underutilized spectrum bands, including the bands that have not been used for cellular communications yet.

Millimeter wave (mmWave) bands from 30 GHz to 300 GHz, previously only considered for outdoor point-to-point backhaul links [9] or for carrying indoor high-resolution multimedia streams [10], have now been put forward as a prime candidate for new spectrum in 5G cellular systems, with the potential bandwidth reaching 10 GHz. This view is supported by recent experiments in New York City that demonstrated the feasibility of mmWave outdoor cellular communications [11], [12]. Originally, the main obstacles for the success of mmWave cellular systems are the huge path loss and rain attenuation, as a result of the ten-fold increase of the carrier frequency [11]. Thanks to the small wavelength of mmWave signals, mmWave MIMO precoding can leverage large-scale antennas at transceivers to provide significant beamforming gains to combat the path loss and to synthesize highly directional beams. Moreover, spectral efficiency can be further increased by transmitting multiple data streams via spatial multiplexing.

For traditional MIMO systems, precoding is typically accomplished at baseband through digital precoders, which can adjust both the magnitude and phase of signals. However, fully digital precoding demands radio frequency (RF) chains, including signal mixers and analog-to-digital converters, comparable in number to the antenna elements. While the small wavelengths of mmWave frequencies facilitate the use of a large number of antenna elements, the prohibitive cost and power consumption of RF chains make digital precoding infeasible. Given such unique constraints in mmWave MIMO systems, a hybrid precoding architecture has recently received much consideration, which only requires a small number of RF chains

1932-4553 © 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

Authorized licensed use limited to: Universite De Limoges. Downloaded on February 13,2020 at 09:07:03 UTC from IEEE Xplore. Restrictions apply.

interfacing between a low-dimensional digital precoder and a high-dimensional analog precoder [13]. As the analog precoders are still of high dimension, it is impractical to implement them in the RF domain with power-hungry variable voltage amplifiers (VGAs) [12]. This heuristic leads to a rule of thumb, i.e., realizing analog precoders with low-cost phase shifters at the expense of sacrificing the ability to change the magnitude of the RF signals.

According to the mapping from RF chains to antennas, which determines the number of phase shifters in use, the hybrid precoding transceiver architectures can be categorized into the fully-connected and partially-connected structures, as illustrated in Fig. 1(b) and Fig. 1(c), respectively. The former structure enjoys full beamforming gain for each RF chain with a natural combination between RF chains and antenna elements, i.e., each RF chain is connected to all antennas. On the other hand, sacrificing some beamforming gain, the partially-connected structure significantly reduces the hardware implementation complexity by connecting each RF chain only with part of the antennas.

In [13], it has been pointed out that maximizing the spectral efficiency of mmWave systems can be approximated by minimizing the Euclidean distance between hybrid precoders and the fully digital precoder. This renders the hybrid precoder design as a matrix factorization problem with unit modulus constraints imposed by the phase shifters. Although significant amounts of research efforts have been invested in solving various matrix factorization problems in recent years [14], [15], with the unique unit modulus constraints, the optimal design of hybrid precoders remains unknown. Existing works often add some extra constraints on analog precoders to simplify the analog part design with unit modulus constraints, which will cause performance loss. This motivates us to reconsider the hybrid precoder design or, in other words, the matrix factorization problem with unit modulus constraints on analog precoders. In particular, a better way to deal with the unit modulus constraint deserves further delicate investigations.

In this paper, by adopting alternating minimization (AltMin) as the main design approach, we will propose different hybrid precoding algorithms to approach the performance of the optimal fully digital precoder. Based on the principle of alternating minimization, three novel algorithms will be proposed to find effective hybrid precoding solutions for the fully-connected and partially-connected structures.

A. Related Works

Hybrid precoding is a newly-emerged technique in mmWave MIMO systems [16]–[20]. So far the main efforts are on the fully-connected structure [13], [21]–[28]. Orthogonal matching pursuit (OMP) is the most widely used algorithm, which often offers reasonably good performance. This algorithm requires the columns of analog precoding matrix to be picked from certain candidate vectors, such as array response vectors of the channel [13], [21], [22], and discrete Fourier transform beamformers [23], [24]. Hence, the OMP-based hybrid precoder design can be viewed as a sparsity constrained

matrix reconstruction problem. Though the design problem is greatly simplified in this way, restricting the space of feasible analog precoding solutions inevitably causes some performance loss. Additionally, extra overhead will be brought up for acquiring the information of array response vectors in advance. More recent attention has mainly focused on reducing the computation complexity of the OMP algorithm [21], [25], e.g., by reusing the matrix inversion result in each iteration.

There are works investigating some special hybrid precoding systems. In [26], an optimal hybrid precoder design in a special case was identified, i.e., when the number of RF chains is at least twice that of the data streams. However, the optimal solution for the general case is unknown. The authors of [28] investigated VGA-enabled hybrid precoding according to different design criteria. By removing VGAs from the RF domain, low-power analog precoders with phase shifters were also considered in [28], whose phases are heuristically extracted from those of the VGA-enabled solution.

On the other hand, much less attention has been paid on the partially-connected structure [29]–[34]. In [29], [30], codebook-based design of hybrid precoders was presented for narrowband and orthogonal frequency division multiplexing (OFDM) systems, respectively. Although the codebook-based design enjoys a low complexity, there will be certain performance loss, and it is not clear how much performance gain can be further obtained. By utilizing the idea of successive interference cancellation (SIC), an iterative hybrid precoding algorithm for the partially-connected structure was proposed in [31]. The algorithm is established based on the assumption that the digital precoding matrix is diagonal, which means that the digital precoder only allocates power to different data streams, and the number of RF chains should be equal to that of the data streams. However, using only analog precoders to provide beamforming gains is obviously a suboptimal strategy [31], [32], which also deviates from the motivation of hybrid precoding. So far there is no study directly optimizing the hybrid precoders without extra constraints in the partially-connected structure, which will be pursued in this paper.

B. Contributions

In this paper, we investigate the hybrid precoder design in mmWave MIMO systems. We will adopt alternating minimization as the main design principle, which helps decouple the precoder design problem into two subproblems, i.e., the analog and digital precoder design. The proposed AltMin algorithms will alternately optimize the digital precoder and the analog precoder. Our major contributions are summarized as follows:

- For the fully-connected structure, we shall show that the unit modulus constraints of the analog precoder define a Riemannian manifold. We will thus propose a manifold optimization based AltMin (MO-AltMin) algorithm. This algorithm does not need any pre-determined candidate set for the analog precoder, and it is the first attempt to directly solve the hybrid precoder design problem under the unit modulus constraints.
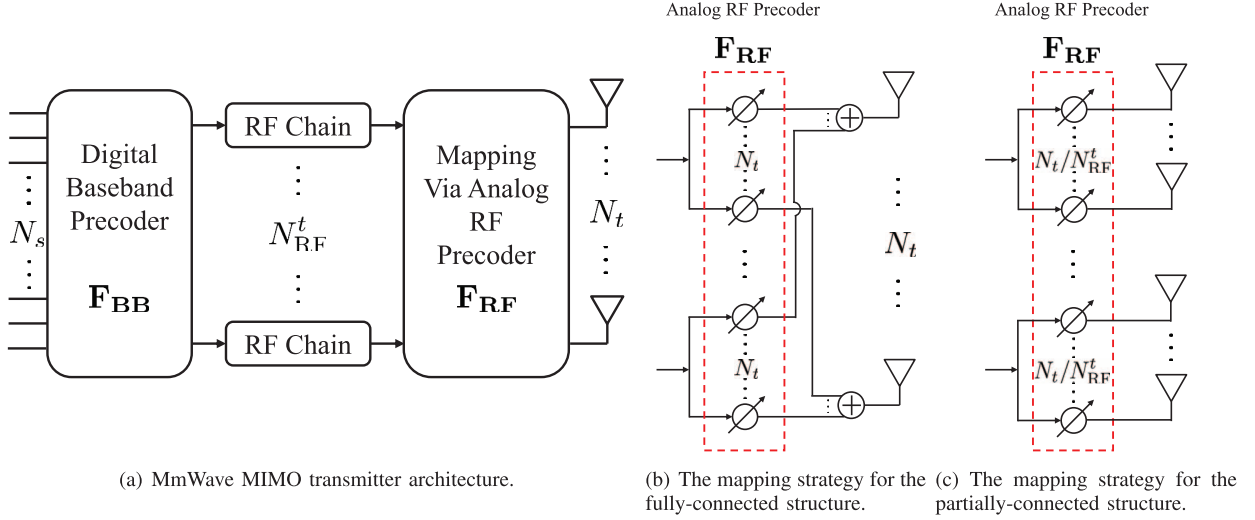
Analog RF Precoder $\mathbf{F_{RF}}$ Analog RF Precoder $\mathbf{F_{RF}}$

(a) MmWave MIMO transmitter architecture.

(b) The mapping strategy for the fully-connected structure.

(c) The mapping strategy for the partially-connected structure.

Fig. 1. Two structures of hybrid precoding in mmWave MIMO systems using different mapping strategies: each RF chain is connected to N_t antennas in (b) and to N_t/N_{RF}^t antennas in (c).

- By imposing an orthogonal property of the digital precoder, we then develop an AltMin algorithm using phase extraction (PE-AltMin) as a low-complexity counterpart of the MO-AltMin algorithm, which will also be more practical for implementation.
- For the partially-connected structure, we propose a semidefinite relaxation based AltMin (SDR-AltMin) algorithm. This algorithm effectively designs the hybrid precoders by offering optimal solutions for both subproblems of analog and digital precoders in each alternating iteration, and it is the first effort directly optimizing the hybrid precoders in such a structure.
- The three proposed AltMin Algorithms can be generally applied to both narrowband and broadband OFDM systems. Simulation results will demonstrate that the MO-AltMin algorithm efficiently identifies a near-optimal solution, while the PE-AltMin algorithm with practical computational complexity outperforms the existing OMP algorithm.
- With the proposed AltMin algorithms, extensive comparisons are provided to reveal valuable design insights. In particular, the proposed AltMin algorithms for the fully-connected structure can help to approach the performance of the fully digital precoder as long as the number of RF chains is comparable to the number of data streams, which cannot be achieved by the widely applied OMP algorithm. On the other hand, the SDR-AltMin algorithm for the partially-connected structure provides significant gains over analog beamforming. Furthermore, by taking advantage of its low-complexity hardware implementation, the partially-connected structure provides a higher energy efficiency than the fully-connected one with a relatively large number of RF chains implemented at transceivers.

Thus, our results firmly establish the effectiveness of the alternating minimization as a key design methodology for hybrid precoder design in mmWave MIMO systems.

C. Organization

The remainder of this paper is organized as follows. We shall introduce the system model and channel model, followed by the problem formulation in Section II. Two alternating minimization algorithms for the fully-connected structure are demonstrated in Section III and Section IV, respectively. In Section V, the hybrid precoder design for the partially-connected structure is investigated. Extensions of the proposed algorithms to OFDM systems are provided in Section VI, and simulation results will be presented in Section VII. Finally, we will conclude this paper in Section VIII.

D. Notations

The following notations are used throughout this paper. $\mathbf{a}$ and $\mathbf{A}$ stand for a column vector and a matrix, respectively; $\mathbf{A}_{i,j}$ is the entry on the ith row and jth column of $\mathbf{A}$; The conjugate, transpose and conjugate transpose of $\mathbf{A}$ are represented by $\mathbf{A}^*$, $\mathbf{A}^T$ and $\mathbf{A}^H$; $\det(\mathbf{A})$ and $\|\mathbf{A}\|_F$ denote the determinant and Frobenius norm of $\mathbf{A}$; $\mathbf{A}^\dagger$ is the Moore-Penrose pseudo inverse of $\mathbf{A}$; $\mathrm{Tr}(\mathbf{A})$ and $\mathrm{vec}(\mathbf{A})$ indicate the trace and vectorization; Expectation and the real part of a complex variable is noted by $\mathbb{E}[\cdot]$ and $\Re[\cdot]$; $\circ$ and $\otimes$ denote the Hadamard and Kronecker products between two matrices.

II. SYSTEM MODEL AND PROBLEM FORMULATION

In this section, we will first present the system model and channel model of the considered mmWave MIMO system, and then formulate the hybrid precoding problem.

A. System Model

Consider a single-user mmWave MIMO system[1] as shown in Fig. 1(a), where N_s data streams are sent and collected by N_t

[1]The receiver side is omitted due to space limitation. More details can be found in [13].

transmit antennas and N_r receive antennas. The numbers of RF chains at the transmitter and receiver are respectively denoted as N_{RF}^t and N_{RF}^r, which are subject to constraints $N_s \leq N_{\mathrm{RF}}^t \leq N_t$ and $N_s \leq N_{\mathrm{RF}}^r \leq N_r$.

The transmitted signal can be written as $\mathbf{x} = \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\mathbf{s}$, where $\mathbf{s}$ is the $N_s \times 1$ symbol vector such that $\mathbb{E}\left[\mathbf{s}\mathbf{s}^H\right] = \frac{1}{N_s}\mathbf{I}_{N_s}$. The hybrid precoders consist of an $N_{\mathrm{RF}}^t \times N_s$ digital baseband precoder $\mathbf{F}_{\mathrm{BB}}$ and an $N_t \times N_{\mathrm{RF}}^t$ analog RF precoder $\mathbf{F}_{\mathrm{RF}}$. The normalized transmit power constraint is given by $\|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_F^2 = N_s$. For simplicity, we first consider a narrow-band block-fading propagation channel, while the extension to broadband OFDM systems will be treated in Section VI. Thus, the received signal after decoding processing is given as

$$\mathbf{y} = \sqrt{\rho}\,\mathbf{W}_{\mathrm{BB}}^H\mathbf{W}_{\mathrm{RF}}^H\mathbf{H}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\mathbf{s} + \mathbf{W}_{\mathrm{BB}}^H\mathbf{W}_{\mathrm{RF}}^H\mathbf{n}, \quad (1)$$

where ρ stands for the average received power, $\mathbf{H}$ is the channel matrix, $\mathbf{W}_{\mathrm{BB}}$ is the $N_{\mathrm{RF}}^r \times N_s$ digital baseband decoder, $\mathbf{W}_{\mathrm{RF}}$ is the $N_r \times N_{\mathrm{RF}}^r$ analog RF decoder at the receiver, and $\mathbf{n}$ is the noise vector of independent and identically distributed (i.i.d.) $\mathcal{CN}(0, \sigma_n^2)$ elements. In this paper, we assume that perfect channel state information (CSI) is known at both the transmitter and receiver. In practice, CSI can be accurately and efficiently obtained by channel estimation [18] at the receiver and further shared at the transmitter with effective feedback techniques [13], [35]. The achievable spectral efficiency when transmitted symbols follow a Gaussian distribution can be expressed as

$$R = \log\det\left(\mathbf{I}_{N_s} + \frac{\rho}{\sigma_n^2 N_s}(\mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}})^\dagger\mathbf{H}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\right.$$
$$\left. \times \mathbf{F}_{\mathrm{BB}}^H\mathbf{F}_{\mathrm{RF}}^H\mathbf{H}^H\left(\mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}\right)\right). \quad (2)$$

Furthermore, the analog precoders are implemented with phase shifters, which can only adjust the phases of the signals. Thus, all the nonzero entries of $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{RF}}$ should satisfy the unit modulus constraints, namely $|(\mathbf{F}_{\mathrm{RF}})_{i,j}| = |(\mathbf{W}_{\mathrm{RF}})_{i,j}| = 1$ for nonzero elements.

According to different signal mapping strategies from RF chains to antennas, the transceiver architecture can be categorized into the fully-connected and partially-connected hybrid precoding structures, as illustrated in Fig. 1(b) and Fig. 1(c). For the fully-connected structure, the output signal of each RF chain is sent to all the antennas through phase shifters, while the partially-connected structure only has N_t/N_{RF}^t antennas connected to each RF chain. Thus, the fully-connected structure enjoys full beamforming gain for each RF chain with a natural combination between RF chains and antennas, whereas the hardware implementation complexity is lower in the partially-connected one by sacrificing some beamforming gain for each RF chain. The structures of $\mathbf{F}_{\mathrm{RF}}$s and $\mathbf{W}_{\mathrm{RF}}$s will vary for different structures, which will be discussed in the following sections in detail.

B. Channel Model

Due to high free-space path loss, the mmWave propagation environment is well characterized by a clustered channel model,

i.e., the Saleh-Valenzuela model [12]. This model depicts the mmWave channel matrix as

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{N_{cl}N_{ray}}}\sum_{i=1}^{N_{cl}}\sum_{l=1}^{N_{ray}}\alpha_{il}\mathbf{a}_r(\phi_{il}^r, \theta_{il}^r)\mathbf{a}_t(\phi_{il}^t, \theta_{il}^t)^H, \quad (3)$$

where N_{cl} and N_{ray} represent the number of clusters and the number of rays in each cluster, and α_{il} denotes the gain of the lth ray in the ith propagation cluster. We assume that α_{il} are i.i.d. random variables following the complex Gaussian distribution $\mathcal{CN}(0, \sigma_{\alpha,i}^2)$, and $\sum_{i=1}^{N_{cl}}\sigma_{\alpha,i}^2 = \hat{\gamma}$, which is the normalization factor to satisfy $\mathbb{E}\left[\|\mathbf{H}\|_F^2\right] = N_t N_r$. In addition, $\mathbf{a}_r(\phi_{il}^r, \theta_{il}^r)$ and $\mathbf{a}_t(\phi_{il}^t, \theta_{il}^t)$ represent the receive and transmit array response vectors, where $\phi_{il}^r(\phi_{il}^t)$ and $\theta_{il}^r(\theta_{il}^t)$ stand for azimuth and elevation angles of arrival and departure (AoAs and AoDs), respectively. In this paper, we consider the uniform square planar array (USPA) with $\sqrt{N} \times \sqrt{N}$ antenna elements. Therefore, the array response vector corresponding to the lth ray in the ith cluster can be written as

$$\mathbf{a}(\phi_{il}, \theta_{il}) = \frac{1}{\sqrt{N}}\left[1, \ldots, e^{j\frac{2\pi}{\lambda}d(p\sin\phi_{il}\sin\theta_{il}+q\cos\theta_{il})},\right.$$
$$\left.\ldots, e^{j\frac{2\pi}{\lambda}d\left(\left(\sqrt{N}-1\right)\sin\phi_{il}\sin\theta_{il}+\left(\sqrt{N}-1\right)\cos\theta_{il}\right)}\right]^T, \quad (4)$$

where d and λ are the antenna spacing and the signal wavelength, and $0 \leq p < \sqrt{N}$ and $0 \leq q < \sqrt{N}$ are the antenna indices in the 2D plane. While this channel model will be used in simulations, our precoder design is applicable to more general models.

C. Problem Formulation

As shown in [13], [21], the design of precoders and decoders can be separated into two subproblems, i.e., the precoding and decoding problems. They have similar mathematical formulations except that there is an extra power constraint in the former. Therefore, we will mainly focus on the precoder design in the remaining part of this paper and the algorithms proposed in this paper can be equally applied for the decoder. The corresponding problem formulation is given by

$$\begin{aligned}\underset{\mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{BB}}}{\text{minimize}} \quad & \|\mathbf{F}_{\mathrm{opt}} - \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_F \\ \text{subject to} \quad & \begin{cases}\mathbf{F}_{\mathrm{RF}} \in \mathcal{A} \\ \|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_F^2 = N_s,\end{cases}\end{aligned} \quad (5)$$

where $\mathbf{F}_{\mathrm{opt}}$ stands for the optimal fully digital precoder, while $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{F}_{\mathrm{BB}}$ are the analog and digital precoders to be optimized. Additionally, $\mathcal{A} \in \{\mathcal{A}_f, \mathcal{A}_p\}$ is the feasible set of the analog precoder induced by the unit modulus constraints, which will be distinct for different hybrid precoding structures.

It has been shown in [13] that minimizing the objective function in (5) approximately leads to the maximization of the spectral efficiency. It is also intuitively true that the optimal hybrid precoders should be sufficiently "close" to the unconstrained optimal digital precoder. In addition, the unconstrained

optimal precoder and decoder are comprised of the first N_s columns of $\mathbf{V}$ and $\mathbf{U}$ respectively, which are unitary matrices derived from the channel's singular value decomposition (SVD), i.e., $\mathbf{H} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H$.

We will mainly treat problem (5) as a matrix factorization problem, for which alternating minimization will be adopted as the main tool. Alternating minimization represents a widely applicable and empirically successful approach for the optimization problems involving different subsets of variables. It has been successfully applied to many applications such as matrix completion [14], phase retrieval [36], image reconstruction [37], blind deconvolution [38] and non-negative matrix factorization [39]. In this paper, the problem formulation (5) is intrinsically a matrix factorization problem involving two matrix variables $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{F}_{\mathrm{BB}}$. However, jointly optimizing these two variables is highly complicated due to the element-wise unit modulus constraints of $\mathbf{F}_{\mathrm{RF}}$. By decoupling the optimization of these two variables, alternating minimization stands out as an efficient method to obtain an effective solution. With the principle of alternating minimization, we will alternately solve for $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{F}_{\mathrm{BB}}$ while fixing the other, which will be the essential idea throughout this paper.

III. MANIFOLD OPTIMIZATION BASED HYBRID PRECODING FOR THE FULLY-CONNECTED STRUCTURE

The fully-connected structure, in which each RF chain is connected to all the antenna elements, is frequently used in mmWave MIMO systems, as shown in Fig. 1(b). This structure restricts every entry in the analog precoding matrix to be unit modulus, and this element-wise constraint makes the precoder design problem intractable. In this section, by observing that the unit modulus constraints define a Riemannian manifold, we will propose an AltMin algorithm based on manifold optimization to directly solve (5).

For the fully-connected structure, inspired by [40], the authors of [26] have shown that the Frobenius norm in (5) can be made exactly zero under the condition that $N_{\mathrm{RF}}^t \geq 2N_s$. This means that the hybrid precoders can achieve the performance of the fully digital precoder in this special case, and the optimal hybrid precoders were obtained in [26]. Thus, we will focus on the region where $N_s \leq N_{\mathrm{RF}}^t < 2N_s$ in this paper.

A. Digital Baseband Precoder Design

We first consider to design the digital precoder $\mathbf{F}_{\mathrm{BB}}$ with a fixed analog precoder $\mathbf{F}_{\mathrm{RF}}$. Thus, problem (5) can be restated as

$$\underset{\mathbf{F}_{\mathrm{BB}}}{\text{minimize}} \quad \|\mathbf{F}_{\mathrm{opt}} - \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_F, \tag{6}$$

which has a well-known least squares solution given by

$$\mathbf{F}_{\mathrm{BB}} = \mathbf{F}_{\mathrm{RF}}^{\dagger}\mathbf{F}_{\mathrm{opt}}. \tag{7}$$

Note that the power constraint in (5) is temporarily removed, and it will be dealt with in Section III-C. Nevertheless, the solution in (7) has already offered a globally optimal solution to the counterpart design problem at the receiver side.
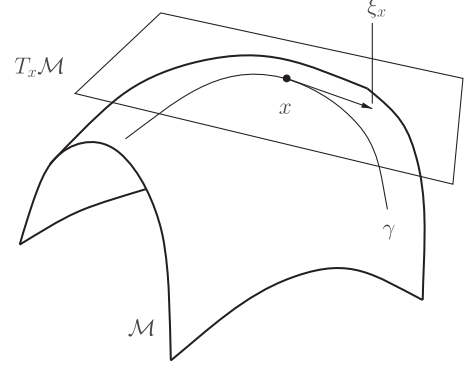

Fig. 2. The tangent space and tangent vector of a Riemannian manifold [41].

B. Analog RF Precoder Design Via Manifold Optimization

For the fully-connected structure, the feasible set $\mathcal{A}_f$ of the analog precoder can be specified by $|(\mathbf{F}_{\mathrm{RF}})_{i,j}| = 1$, as each RF chain is connected to all the antennas. In the next alternating step, we fix $\mathbf{F}_{\mathrm{BB}}$ and seek an analog precoder which optimizes the following problem[2]:

$$\begin{aligned} \underset{\mathbf{F}_{\mathrm{RF}}}{\text{minimize}} \quad & \|\mathbf{F}_{\mathrm{opt}} - \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_F^2 \\ \text{subject to} \quad & |(\mathbf{F}_{\mathrm{RF}})_{i,j}| = 1, \forall i, j. \end{aligned} \tag{8}$$

The main obstacles are the unit modulus constraints, which are intrinsically non-convex. To the best of the authors' knowledge, there is no general approach to solve (8) optimally. In the following, we will propose an effective manifold optimization algorithm to find a near-optimal solution of problem (8).

We will start with some definitions and terminologies in manifold optimization. More background on manifolds and manifold optimization can be found in [41]–[43], and there are some recent applications in wireless communications [44]. As shown in Fig. 2, a *manifold* $\mathcal{M}$ is a topological space that resembles a Euclidean space near each point [43]. In other words, each point on a manifold has a neighborhood that is homeomorphic to the Euclidean space. The *tangent space* $T_x\mathcal{M}$ at a given point x on the manifold $\mathcal{M}$ is composed of the *tangent vectors* ξ_x of the curves γ through the point x. In most applications, manifolds fall into a special category of topological manifold, namely, a *Riemannian manifold*. A Riemannian manifold is equipped with an inner product defined on the tangent spaces $T_x\mathcal{M}$, called the *Riemannian metric*, which allows one to measure distances and angles on manifolds. In particular, it is possible to use calculus on a Riemannian manifold with the Riemannian metric.

The rich geometry of Riemannian manifolds makes it possible to define gradients of cost functions. More importantly, optimization over a Riemannian manifold is locally analogous to that over a Euclidean space with smooth constraints. Therefore, a well-developed conjugate gradient algorithm in Euclidean spaces can find its counterpart on the specified Riemannian manifolds. In the following, we will briefly introduce this counterpart.

[2]The square of the Frobenius norm makes the objective function quadratic and smooth, and will not affect the solution.

We first endow the complex plane $\mathbb{C}$ with the Euclidean metric

$$\langle x_1, x_2 \rangle = \Re\{x_1^* x_2\}, \tag{9}$$

which is equivalent to treating $\mathbb{C}$ as $\mathbb{R}^2$ with the canonical inner product. Then we are able to denote the complex circle as

$$\mathcal{M}_{cc} = \{x \in \mathbb{C} : x^* x = 1\}. \tag{10}$$

For a given point x on the manifold $\mathcal{M}_{cc}$, the directions along which it can move are characterized by the tangent vectors. Then the tangent space at the point $x \in \mathcal{M}_{cc}$ can be represented by

$$T_x \mathcal{M}_{cc} = \{z \in \mathbb{C} : z^* x + x^* z = 2\langle x, z \rangle = 0\}. \tag{11}$$

Note that the vector $\mathbf{x} = \text{vec}(\mathbf{F}_{\text{RF}})$ forms a complex circle manifold $\mathcal{M}_{cc}^m = \{\mathbf{x} \in \mathbb{C}^m : |\mathbf{x}_1| = |\mathbf{x}_2| = \cdots = |\mathbf{x}_m| = 1\}$, where $m = N_t N_{\text{RF}}^t$. Therefore, the search space of the optimization problem (8) is over a product of m circles in the complex plane, which is a Riemannian submanifold of $\mathbb{C}^m$ with the product geometry. Hence, the tangent space at a given point $\mathbf{x} \in \mathcal{M}_{cc}^m$ can be expressed as

$$T_{\mathbf{x}} \mathcal{M}_{cc}^m = \{\mathbf{z} \in \mathbb{C}^m : \Re\{\mathbf{z} \circ \mathbf{x}^*\} = \mathbf{0}_m\}. \tag{12}$$

Among all the tangent vectors, similar to the Euclidean space, one of them that is related to the negative *Riemannian gradient* represents the direction of the greatest decrease of a function. Because the complex circle manifold $\mathcal{M}_{cc}^m$ is a Riemannian submanifold of $\mathbb{C}^m$, the Riemannian gradient at $\mathbf{x}$ is a tangent vector $\text{grad} f(\mathbf{x})$ given by the orthogonal projection of the Euclidean gradient $\nabla f(\mathbf{x})$ onto the tangent space $T_{\mathbf{x}} \mathcal{M}_{cc}^m$ [41]:

$$\begin{aligned} \text{grad} f(\mathbf{x}) &= \text{Proj}_{\mathbf{x}} \nabla f(\mathbf{x}) \\ &= \nabla f(\mathbf{x}) - \Re\{\nabla f(\mathbf{x}) \circ \mathbf{x}^*\} \circ \mathbf{x}, \end{aligned} \tag{13}$$

where the Euclidean gradient of the cost function in (8) is

$$\nabla f(\mathbf{x}) = -2(\mathbf{F}_{\text{BB}}^* \otimes \mathbf{I}_{N_t})\left[\text{vec}(\mathbf{F}_{\text{opt}}) - (\mathbf{F}_{\text{BB}}^T \otimes \mathbf{I}_{N_t})\mathbf{x}\right]. \tag{14}$$

Solving this Euclidean gradient involves some techniques on complex-valued matrix derivatives, the details can be found in [45].

Retraction is another key factor in manifold optimization, which maps a vector from the tangent space onto the manifold itself. It determines the destination on the manifold when moving along a tangent vector. The retraction of a tangent vector $\alpha \mathbf{d}$ at point $\mathbf{x} \in \mathcal{M}_{cc}^m$ can be stated as

$$\begin{aligned} \text{Retr}_{\mathbf{x}} : T_{\mathbf{x}} \mathcal{M}_{cc}^m &\to \mathcal{M}_{cc}^m : \\ \alpha \mathbf{d} \mapsto \text{Retr}_{\mathbf{x}}(\alpha \mathbf{d}) &= \text{vec}\left[\frac{(\mathbf{x} + \alpha \mathbf{d})_i}{|(\mathbf{x} + \alpha \mathbf{d})_i|}\right]. \end{aligned} \tag{15}$$

Equipped with the tangent space, Riemannian gradient and retraction of the complex circle manifold $\mathcal{M}_{cc}^m$, a line search based conjugate gradient method [46], which is a classical algorithm in the Euclidean space, can be developed to design the analog precoder as shown in Algorithm 1.

Algorithm 1 Conjugate Gradient Algorithm for Analog Precoding Based on Manifold Optimization

Input: $\mathbf{F}_{\text{opt}}, \mathbf{F}_{\text{BB}}, \mathbf{x}_0 \in \mathcal{M}_{cc}^m$
1: $\mathbf{d}_0 = -\text{grad} f(\mathbf{x}_0)$ and $k = 0$;
2: **repeat**
3:      Choose Armijo backtracking line search step size α_k;
4:      Find the next point $\mathbf{x}_{k+1}$ using retraction in (15): $\mathbf{x}_{k+1} = \text{Retr}_{\mathbf{x}_k}(\alpha_k \mathbf{d}_k)$;
5:      Determine Riemannian gradient $\mathbf{g}_{k+1} = \text{grad} f(\mathbf{x}_{k+1})$ according to (13) and (14);
6:      Calculate the vector transports $\mathbf{g}_k^+$ and $\mathbf{d}_k^+$ of gradient $\mathbf{g}_k$ and conjugate direction $\mathbf{d}_k$ from $\mathbf{x}_k$ to $\mathbf{x}_{k+1}$;
7:      Choose Polak-Ribiere parameter β_{k+1};
8:      Compute conjugate direction $\mathbf{d}_{k+1} = -\mathbf{g}_{k+1} + \beta_{k+1} \mathbf{d}_k^+$;
9:      $k \leftarrow k + 1$;
10: **until** a stopping criterion triggers.

Algorithm 1 utilizes the well-known Armijo backtracking line search step and Polak-Ribiere parameter to guarantee the objective function to be non-increasing in each iteration [47]. In addition, since Steps (7) and (8) involve the operations between two vectors in different tangent spaces $T_{\mathbf{x}_k} \mathcal{M}_{cc}^m$ and $T_{\mathbf{x}_{k+1}} \mathcal{M}_{cc}^m$, which cannot be combined directly, a mapping between two tangent vectors in different tangent spaces called *transport* is introduced. The transport of a tangent vector $\mathbf{d}$ from $\mathbf{x}_k$ to $\mathbf{x}_{k+1}$ can be specified as

$$\begin{aligned} \text{Transp}_{\mathbf{x}_k \to \mathbf{x}_{k+1}} : T_{\mathbf{x}_k} \mathcal{M}_{cc}^m &\to T_{\mathbf{x}_{k+1}} \mathcal{M}_{cc}^m : \\ \mathbf{d} \mapsto \mathbf{d} - \Re\{\mathbf{d} \circ \mathbf{x}_{k+1}^*\} \circ \mathbf{x}_{k+1}, \end{aligned} \tag{16}$$

which is accomplished in Step 6. According to Theorem 4.3.1 in [41], Algorithm 1 is guaranteed to converge to a critical point, i.e., the point where the gradient of the objective function is zero.

C. Hybrid Precoder Design

With Algorithm 1 at hand, the hybrid precoder design via alternating minimization for the fully-connected structure is described in the **MO-AltMin Algorithm** by solving problems (6) and (8) iteratively. To satisfy the power constraint in (5), we normalize $\mathbf{F}_{\text{BB}}$ by a factor of $\frac{\sqrt{N_s}}{\|\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\|_F}$ at Step (7). The following lemma help reveal the effect of this normalization.

MO-AltMin Algorithm: Manifold Optimization Based Hybrid Precoding for the Fully-connected Structure

Input: $\mathbf{F}_{\text{opt}}$
1: Construct $\mathbf{F}_{\text{RF}}^{(0)}$ with random phases and set $k = 0$;
2: **repeat**
3:      Fix $\mathbf{F}_{\text{RF}}^{(k)}$, and $\mathbf{F}_{\text{BB}}^{(k)} = \mathbf{F}_{\text{RF}}^{(k)\dagger} \mathbf{F}_{\text{opt}}$;
4:      Optimize $\mathbf{F}_{\text{RF}}^{(k+1)}$ using Algorithm 1 when $\mathbf{F}_{\text{BB}}^{(k)}$ is fixed;
5:      $k \leftarrow k + 1$;
6: **until** a stopping criterion triggers;
7: For the digital precoder at the transmit end, normalize $\widehat{\mathbf{F}}_{\text{BB}} = \frac{\sqrt{N_s}}{\|\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\|_F} \mathbf{F}_{\text{BB}}$.

Lemma 1: If the Euclidean distance before normalization is $\|\mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\|_F \leq \delta$, then after normalization we have $\left\|\mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{RF}}\widehat{\mathbf{F}}_{\text{BB}}\right\|_F \leq 2\delta$.

Proof: Define the normalization factor $\frac{\sqrt{N_s}}{\|\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\|_F} = \frac{1}{\lambda}$ and thus $\|\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\|_F = \lambda\sqrt{N_s} = \lambda\|\mathbf{F}_{\text{opt}}\|_F$.

By norm inequality, we have

$$
\begin{aligned}
\|\mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\|_F &\geq \big|\,\|\mathbf{F}_{\text{opt}}\|_F - \|\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\|_F\,\big| \\
&= |1 - \lambda|\,\|\mathbf{F}_{\text{opt}}\|_F, \quad\quad (17)
\end{aligned}
$$

which is equivalent to $\|\mathbf{F}_{\text{opt}}\|_F \leq \frac{1}{|\lambda-1|}\delta$.

When $\lambda \neq 1$, which indicates $\|\mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\|_F \neq 0$,

$$
\begin{aligned}
&\left\|\mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{RF}}\widehat{\mathbf{F}}_{\text{BB}}\right\|_F \\
&= \left\|\mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}} + \left(1 - \frac{1}{\lambda}\right)\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\right\|_F \\
&\leq \|\mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\|_F + \left|1 - \frac{1}{\lambda}\right|\|\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\|_F \\
&\leq \delta + |\lambda-1|\,\|\mathbf{F}_{\text{opt}}\|_F \leq \delta + \frac{|\lambda-1|}{|\lambda-1|}\delta = 2\delta. \quad (18)
\end{aligned}
$$
∎

Lemma 1 shows that as long as we can make the Euclidean distance between the optimal digital precoder and the hybrid precoders sufficiently small when ignoring the power constraint in (5), the normalization step will also achieve a small distance to the optimal digital precoder.

Since the objective function in problem (5) is minimized at Steps 3 and 4, each iteration will never increase it. In addition, the objective function is non-negative. These two properties together guarantee that the MO-AltMin algorithm can converge to a feasible solution. Although the optimality of alternating minimization algorithms for general non-convex problems is still an open problem [48], simulation results in Section VII will show that the proposed algorithm can provide near-optimal performance.

However, the complexity of the MO-AltMin algorithm is relatively high. In each iteration, the update of the analog precoder involves a line search algorithm, i.e., Algorithm 1, so the nested loops in the MO-AltMin algorithm will slow down the whole solving procedure. Furthermore, the Kronecker products in (14) will result in two matrices of dimension $N_{\text{RF}}^t N_t \times N_s N_t$, which scales with the antenna size and results in an exponential increase of the computational complexity in the MO-AltMin algorithm. Despite the high complexity, we note that the MO-AltMin algorithm based on manifold optimization directly solves the hybrid precoder design problem (5) under unit modulus constraints, which will improve the spectral efficiency when compared to existing algorithms. Therefore, this algorithm can serve as a benchmark of the performance in terms of spectral efficiency, and we will seek a low-complexity algorithm in the next section.

IV. LOW-COMPLEXITY HYBRID PRECODING FOR THE FULLY-CONNECTED STRUCTURE

Although the MO-AltMin algorithm can directly handle the unit modulus constraints, the number of such constraints may be substantially large due to the large-size antenna array. Thus, the high computational complexity will prevent its practical implementation. It motivates us to develop a hybrid precoding algorithm with lower computational complexity and slight performance loss. In this section, by utilizing the orthogonal property of the digital precoder, we will propose a low-complexity design for the analog precoder subject to unit modulus constraints. Thanks to the orthogonal property of the digital precoder, the phases of the analog precoder can be extracted from the phases of an equivalent precoder determined by the digital precoder and the unconstrained optimal digital precoder. Though it will incur some performance loss compared to the MO-AltMin algorithm, simulations will demonstrate its performance gains over existing algorithms.

A. Digital Baseband Precoder Structure

Note that the columns of the unconstrained optimal precoding matrix $\mathbf{F}_{\text{opt}}$ are mutually orthogonal in order to mitigate the interference between the multiplexed streams. Inspired by this structure of the unconstrained precoding solution, we impose a similar constraint that the columns of the digital precoding matrix should be mutually orthogonal, i.e.,

$$
\mathbf{F}_{\text{BB}}^H\mathbf{F}_{\text{BB}} = \alpha\mathbf{F}_{\text{DD}}^H\alpha\mathbf{F}_{\text{DD}} = \alpha^2\mathbf{I}_{N_s}, \quad (19)
$$

where $\mathbf{F}_{\text{DD}}$ is a unitary matrix with the same dimension as $\mathbf{F}_{\text{BB}}$. Although there is no existing conclusion on the optimal structure of the digital precoder in hybrid precoding, it is natural and intriguing to investigate the hybrid precoder design under such an orthogonal constraint of the digital precoder. More importantly, this orthogonal constraint creates the potential for the analog precoder $\mathbf{F}_{\text{RF}}$ to get rid of the product form with $\mathbf{F}_{\text{BB}}$, which will help significantly simplify the analog precoder design.

B. Hybrid Precoder Design

By replacing $\mathbf{F}_{\text{BB}}$ with $\alpha\mathbf{F}_{\text{DD}}$, the objective function in (5) can be further recast as

$$
\begin{aligned}
&\|\mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\|_F^2 \\
&= \text{Tr}\left(\mathbf{F}_{\text{opt}}^H\mathbf{F}_{\text{opt}}\right) - \text{Tr}\left(\mathbf{F}_{\text{opt}}^H\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\right) \\
&\quad - \text{Tr}\left(\mathbf{F}_{\text{BB}}^H\mathbf{F}_{\text{RF}}^H\mathbf{F}_{\text{opt}}\right) + \text{Tr}\left(\mathbf{F}_{\text{BB}}^H\mathbf{F}_{\text{RF}}^H\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\right) \\
&= \|\mathbf{F}_{\text{opt}}\|_F^2 - 2\alpha\Re\text{Tr}\left(\mathbf{F}_{\text{DD}}\mathbf{F}_{\text{opt}}^H\mathbf{F}_{\text{RF}}\right) \\
&\quad + \alpha^2\|\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{DD}}\|_F^2. \quad (20)
\end{aligned}
$$

Obviously, when $\alpha = \frac{\Re\text{Tr}\left(\mathbf{F}_{\text{DD}}\mathbf{F}_{\text{opt}}^H\mathbf{F}_{\text{RF}}\right)}{\|\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{DD}}\|_F^2}$, the objective function $\|\mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\|_F^2$ in (20) has the minimum value, given by $\|\mathbf{F}_{\text{opt}}\|_F^2 - \frac{\left\{\Re\text{Tr}\left(\mathbf{F}_{\text{DD}}\mathbf{F}_{\text{opt}}^H\mathbf{F}_{\text{RF}}\right)\right\}^2}{\|\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{DD}}\|_F^2}$. Note that the square of the

Frobenius norm $\|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{DD}}\|_F^2$ has the following upper bound

$$
\begin{aligned}
\|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{DD}}\|_F^2 &= \mathrm{Tr}\left(\mathbf{F}_{\mathrm{DD}}^H\mathbf{F}_{\mathrm{RF}}^H\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{DD}}\right) \\
&= \mathrm{Tr}\left\{\begin{pmatrix}\mathbf{I}_{N_s} \\ \mathbf{0}\end{pmatrix}\mathbf{K}^H\mathbf{F}_{\mathrm{RF}}^H\mathbf{F}_{\mathrm{RF}}\mathbf{K}\right\} \\
&\leq \mathrm{Tr}\left\{\mathbf{K}^H\mathbf{F}_{\mathrm{RF}}^H\mathbf{F}_{\mathrm{RF}}\mathbf{K}\right\} \\
&= \|\mathbf{F}_{\mathrm{RF}}\|_F^2,
\end{aligned}
\tag{21}
$$

where $\mathbf{F}_{\mathrm{DD}}\mathbf{F}_{\mathrm{DD}}^H = \mathbf{K}\begin{pmatrix}\mathbf{I}_{N_s} \\ \mathbf{0}\end{pmatrix}\mathbf{K}^H$ is the SVD of $\mathbf{F}_{\mathrm{DD}}\mathbf{F}_{\mathrm{DD}}^H$ and the equality holds when $N_{\mathrm{RF}}^t = N_s$, i.e., $\mathbf{F}_{\mathrm{DD}}$ is a square matrix. Hence, the objective function in (5) is upper bounded by $\|\mathbf{F}_{\mathrm{opt}}\|_F^2 - \frac{\{\Re\mathrm{Tr}(\mathbf{F}_{\mathrm{DD}}\mathbf{F}_{\mathrm{opt}}^H\mathbf{F}_{\mathrm{RF}})\}^2}{\|\mathbf{F}_{\mathrm{RF}}\|_F^2}$. In order to make $\mathbf{F}_{\mathrm{RF}}$ get rid of the product with $\mathbf{F}_{\mathrm{BB}}$, we choose to add the constant term $\left(\frac{1}{2\|\mathbf{F}_{\mathrm{RF}}\|_F^2} - 1\right)\|\mathbf{F}_{\mathrm{opt}}\|_F^2 + \frac{1}{2}$ to the bound and multiply it by the positive constant term $2\|\mathbf{F}_{\mathrm{RF}}\|_F^2$. Then we have

$$
\begin{aligned}
&\|\mathbf{F}_{\mathrm{opt}}\|_F^2 - 2\Re\mathrm{Tr}\left(\mathbf{F}_{\mathrm{DD}}\mathbf{F}_{\mathrm{opt}}^H\mathbf{F}_{\mathrm{RF}}\right) + \|\mathbf{F}_{\mathrm{RF}}\|_F^2 \\
&= \mathrm{Tr}\left(\mathbf{F}_{\mathrm{RF}}^H\mathbf{F}_{\mathrm{RF}}\right) - 2\Re\mathrm{Tr}\left(\mathbf{F}_{\mathrm{DD}}\mathbf{F}_{\mathrm{opt}}^H\mathbf{F}_{\mathrm{RF}}\right) \\
&\quad + \mathrm{Tr}\left(\mathbf{F}_{\mathrm{DD}}\mathbf{F}_{\mathrm{opt}}^H\mathbf{F}_{\mathrm{opt}}\mathbf{F}_{\mathrm{DD}}^H\right) \\
&= \left\|\mathbf{F}_{\mathrm{opt}}\mathbf{F}_{\mathrm{DD}}^H - \mathbf{F}_{\mathrm{RF}}\right\|_F^2.
\end{aligned}
\tag{22}
$$

Since directly optimizing the objective function (20) will still incur high complexity, we intend to adopt the upper bound (22) as the objective function rather than the original one. In addition, as we can satisfy the transmit power constraint by normalization after updating the hybrid precoders alternately, which has been shown in the MO-AltMin algorithm and Lemma 1, here we also temporarily remove the power constraint. Thus, by adopting (22) as the objective function, the hybrid precoder design problem is given as

$$
\begin{aligned}
&\underset{\mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{DD}}}{\mathrm{minimize}} \quad \left\|\mathbf{F}_{\mathrm{opt}}\mathbf{F}_{\mathrm{DD}}^H - \mathbf{F}_{\mathrm{RF}}\right\|_F^2 \\
&\text{subject to} \quad \begin{cases}|(\mathbf{F}_{\mathrm{RF}})_{i,j}| = 1, \forall i,j \\ \mathbf{F}_{\mathrm{DD}}^H\mathbf{F}_{\mathrm{DD}} = \mathbf{I}_{N_s}.\end{cases}
\end{aligned}
\tag{23}
$$

The problem formulation (23) implies that we only need to seek a unitary precoding matrix $\mathbf{F}_{\mathrm{DD}}$, and then a corresponding precoding matrix $\mathbf{F}_{\mathrm{BB}}$ with orthogonal columns can be obtained. When applying alternating minimization, it turns out that the objective function in (23) significantly simplifies the analog precoder design. More specifically, since the matrix $\mathbf{F}_{\mathrm{RF}}$ gets rid of the product form with $\mathbf{F}_{\mathrm{BB}}$, it has a closed-form solution

$$
\arg\left(\mathbf{F}_{\mathrm{RF}}\right) = \arg\left(\mathbf{F}_{\mathrm{opt}}\mathbf{F}_{\mathrm{DD}}^H\right),
\tag{24}
$$

where $\arg(\mathbf{A})$ generates a matrix containing the phases of the entries of $\mathbf{A}$. Thus, it shows that the phases of $\mathbf{F}_{\mathrm{RF}}$ can be extracted from the phases of an equivalent precoder $\mathbf{F}_{\mathrm{opt}}\mathbf{F}_{\mathrm{DD}}^H$. This closed-form solution can also be viewed as the Euclidean projection of $\mathbf{F}_{\mathrm{opt}}\mathbf{F}_{\mathrm{DD}}^H$ on the feasible set $\mathcal{A}_f$ of the analog precoder.

For the digital precoder design, regarding $\mathbf{F}_{\mathrm{RF}}$ as fixed, we try to solve a digital precoder which optimizes the following problem

$$
\begin{aligned}
&\underset{\mathbf{F}_{\mathrm{DD}}}{\mathrm{minimize}} \quad \left\|\mathbf{F}_{\mathrm{opt}}\mathbf{F}_{\mathrm{DD}}^H - \mathbf{F}_{\mathrm{RF}}\right\|_F^2 \\
&\text{subject to} \quad \mathbf{F}_{\mathrm{DD}}^H\mathbf{F}_{\mathrm{DD}} = \mathbf{I}_{N_s}.
\end{aligned}
\tag{25}
$$

Since problem (25) only has one optimization variable $\mathbf{F}_{\mathrm{DD}}$, it is equivalent to

$$
\begin{aligned}
&\underset{\mathbf{F}_{\mathrm{DD}}}{\mathrm{maximize}} \quad \Re\mathrm{Tr}\left(\mathbf{F}_{\mathrm{DD}}\mathbf{F}_{\mathrm{opt}}^H\mathbf{F}_{\mathrm{RF}}\right) \\
&\text{subject to} \quad \mathbf{F}_{\mathrm{DD}}^H\mathbf{F}_{\mathrm{DD}} = \mathbf{I}_{N_s}.
\end{aligned}
\tag{26}
$$

According to the definition of the dual norm, we have

$$
\begin{aligned}
\Re\mathrm{Tr}\left(\mathbf{F}_{\mathrm{DD}}\mathbf{F}_{\mathrm{opt}}^H\mathbf{F}_{\mathrm{RF}}\right) &\leq \left|\mathrm{Tr}\left(\mathbf{F}_{\mathrm{DD}}\mathbf{F}_{\mathrm{opt}}^H\mathbf{F}_{\mathrm{RF}}\right)\right| \\
&\overset{(a)}{\leq} \left\|\mathbf{F}_{\mathrm{DD}}^H\right\|_\infty \cdot \left\|\mathbf{F}_{\mathrm{opt}}^H\mathbf{F}_{\mathrm{RF}}\right\|_1 \\
&= \left\|\mathbf{F}_{\mathrm{opt}}^H\mathbf{F}_{\mathrm{RF}}\right\|_1 \\
&= \sum_{i=1}^{N_s}\sigma_i,
\end{aligned}
\tag{27}
$$

where (a) follows the Hölder's inequality, $\|\cdot\|_\infty$ and $\|\cdot\|_1$ stand for the infinite and one Schatten norms [49]. The equality is established only when

$$
\mathbf{F}_{\mathrm{DD}} = \mathbf{V}_1\mathbf{U}^H,
\tag{28}
$$

where $\mathbf{F}_{\mathrm{opt}}^H\mathbf{F}_{\mathrm{RF}} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H = \mathbf{U}\mathbf{S}\mathbf{V}_1^H$, which is the SVD of $\mathbf{F}_{\mathrm{opt}}^H\mathbf{F}_{\mathrm{RF}}$, and $\mathbf{S}$ is a diagonal matrix whose elements are the first N_s nonzero singular values $\sigma_1, \ldots, \sigma_{N_s}$. This result bears some similarity to the solution of the orthogonal Procrustes problem (OPP) [50], although the formulation is slightly different[3].

PE-AltMin Algorithm: A Low-Complexity Algorithm for the Fully-connected Structure

---

Input: $\mathbf{F}_{\mathrm{opt}}$
1: Construct $\mathbf{F}_{\mathrm{RF}}^{(0)}$ with random phases and set $k = 0$;
2: **repeat**
3:     Fix $\mathbf{F}_{\mathrm{RF}}^{(k)}$, compute the SVD: $\mathbf{F}_{\mathrm{opt}}^H\mathbf{F}_{\mathrm{RF}}^{(k)}=\mathbf{U}^{(k)}\mathbf{S}^{(k)}\mathbf{V}_1^{(k)H}$;
4:     $\mathbf{F}_{\mathrm{DD}}^{(k)} = \mathbf{V}_1^{(k)}\mathbf{U}^{(k)H}$;
5:     Fix $\mathbf{F}_{\mathrm{DD}}^{(k)}$, and $\arg\left\{\mathbf{F}_{\mathrm{RF}}^{(k+1)}\right\} = \arg\left(\mathbf{F}_{\mathrm{opt}}\mathbf{F}_{\mathrm{DD}}^{(k)H}\right)$;
6:     $k \leftarrow k + 1$;
7: **until** a stopping criterion triggers;
8: For the digital precoder at the transmit end, normalize $\widehat{\mathbf{F}}_{\mathrm{BB}} = \frac{\sqrt{N_s}}{\|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{DD}}\|_F}\mathbf{F}_{\mathrm{DD}}$.

Based on the two closed-form solutions for the analog and digital precoders, we summarize our new design as the **PE-AltMin Algorithm**. There are several issues involved in the PE-AltMin algorithm that require some further remarks.

---

[3]OPP tries to minimize $\|\mathbf{A}\boldsymbol{\Omega} - \mathbf{B}\|_F$, where the optimization variable $\boldsymbol{\Omega}$ is a square unitary matrix.

(1) Complexity: In both the MO-AltMin and PE-AltMin algorithms, the updating rules of the digital precoders are given by closed-form solutions and thus these two algorithms are of comparable complexity in the digital parts. Furthermore, in the hybrid precoding system, the dimension of the analog precoder is much higher than that of the digital precoder, which makes the complexity of the algorithms predominated by the analog part.

In each iteration of the MO-AltMin algorithm, a conjugate gradient descent search is needed to update the analog precoder. In particular, when updating the analog precoder in each iteration, we need to search on the complex circle manifold repeatedly to find a local optimum with zero gradient of the cost function. Additionally, the computation of the large-size matrices, i.e., matrices of dimension $N_{\mathrm{RF}}^t N_t \times N_s N_t$, will be involved in the gradient descent procedure due to the use of the Kronecker products. More importantly, the conjugate gradient descent, which is an iterative procedure itself, is nested into each alternating minimization iteration. This nested iteration structure will dramatically degrade the computational efficiency of the MO-AltMin algorithm. On the contrary, in each iteration of the PE-AltMin algorithm, the update of the analog precoder is simply realized by a phase extraction operation of the matrix $\mathbf{F}_{\mathrm{opt}}\mathbf{F}_{\mathrm{DD}}^H$, whose dimension is $N_t \times N_{\mathrm{RF}}^t$. Compared with the MO-AltMin algorithm, it is safe to conclude that the PE-AltMin algorithm is with a much lower complexity, which is also numerically observed in our simulations.

(2) Accuracy of the approximation: In the formulation of (23), we try to minimize an upper bound given by (21) rather than directly minimizing the original objective function. Therefore, the effectiveness of this strategy depends on how tight the upper bound is when $N_s \leq N_{\mathrm{RF}}^t < 2N_s$. According to (21), we can quantify the gap between $\|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{DD}}\|_F^2$ and $\|\mathbf{F}_{\mathrm{RF}}\|_F^2$ as $\|\mathbf{F}_{\mathrm{RF}}\mathbf{K}_1\|_F^2$, where $\mathbf{K}_1$ is comprised of the rightmost $N_{\mathrm{RF}}^t - N_s$ columns of $\mathbf{K}$. Note that when $N_{\mathrm{RF}}^t = N_s$, the upper bound is tight, i.e., the equality holds in (21). Furthermore, when increasing N_{RF}^t from N_s to $2N_s - 1$, the gap $\|\mathbf{F}_{\mathrm{RF}}\mathbf{K}_1\|_F^2$ will become larger with increasing N_{RF}^t, which will result in some performance loss. The impact of adopting this upper bound as the objective will be shown in Section VII via simulations.

(3) Calculation of α: Although we have stated that a value of α can be found to construct a matrix $\mathbf{F}_{\mathrm{BB}}$ corresponding to each $\mathbf{F}_{\mathrm{DD}}$, we do not need to actually calculate α in the PE-AltMin algorithm for the following reasons. At the transmitter side, even though we compute $\mathbf{F}_{\mathrm{BB}} = \alpha\mathbf{F}_{\mathrm{DD}}$ in the last step of the PE-AltMin algorithm, this digital precoder should be immediately normalized. Hence, the overall procedure is equivalent to directly normalizing $\mathbf{F}_{\mathrm{DD}}$ to satisfy the power constraint without knowing α. At the receiver side, we note that the spectral efficiency (2) will not be influenced by the constant factor α multiplied with $\mathbf{W}_{\mathrm{BB}}$. That is because the decoder $\mathbf{W}_{\mathrm{BB}}$ takes effects on both received signals and noise, and thus signal-to-noise ratio (SNR) will not change due to the constant factor α. Moreover, the avoidance of calculating α will further reduce the complexity of the PE-AltMin algorithm.

V. HYBRID PRECODING FOR THE PARTIALLY CONNECTED STRUCTURE

Different from the fully-connected structure, the partially-connected structure shown in Fig. 1(c) [29], [33], [34], also called the *array of subarray structure*, employs notably less phase shifters and is advocated for energy-efficient mmWave MIMO systems [31], [32]. Particularly, the output signal of each RF chain is only connected with N_t/N_{RF}^t antennas, which reduces the hardware complexity in the RF domain. Therefore, the analog precoder $\mathbf{F}_{\mathrm{RF}}$ in this structure belongs to a set of block matrices $\mathcal{A}_p$, where each block is an N_t/N_{RF}^t dimension vector with unit modulus elements and the structure of $\mathbf{F}_{\mathrm{RF}}$ can be depicted as

$$\mathbf{F}_{\mathrm{RF}} = \begin{bmatrix} \mathbf{p}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{p}_2 & & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{p}_{N_{\mathrm{RF}}^t} \end{bmatrix}, \qquad (29)$$

where $\mathbf{p}_i = \left[\exp\left(j\theta_{(i-1)\frac{N_t}{N_{\mathrm{RF}}^t}+1}\right), \ldots, \exp\left(j\theta_{i\frac{N_t}{N_{\mathrm{RF}}^t}}\right)\right]^T$ and θ_i stands for the phase of the ith phase shifter. In this section, we will propose an AltMin algorithm for this structure. Surprisingly, optimal solutions can be found for both subproblems of analog and digital precoders.

A. Analog RF Precoder Design

Due to the special structure of the constraint on $\mathbf{F}_{\mathrm{RF}}$, in the product $\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}$, each nonzero element of $\mathbf{F}_{\mathrm{RF}}$ is multiplied by a corresponding row extracted from $\mathbf{F}_{\mathrm{BB}}$. Thus, the power constraint in (5) at the transmit side can be recast as

$$\|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_F^2 = \frac{N_t}{N_{\mathrm{RF}}^t}\|\mathbf{F}_{\mathrm{BB}}\|_F^2 = N_s. \qquad (30)$$

Therefore, the analog precoder design is formulated as

$$\begin{aligned} \underset{\mathbf{F}_{\mathrm{RF}}}{\mathrm{minimize}} \quad & \|\mathbf{F}_{\mathrm{opt}} - \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_F^2 \\ \mathrm{subject\ to} \quad & \mathbf{F}_{\mathrm{RF}} \in \mathcal{A}_p. \end{aligned} \qquad (31)$$

Also, due to the same property of $\mathbf{F}_{\mathrm{RF}}$, problem (31) can be reformulated as

$$\underset{\{\theta_i\}_{i=1}^{N_t}}{\mathrm{minimize}} \quad \left\|(\mathbf{F}_{\mathrm{opt}})_{i,:} - e^{j\theta_i}(\mathbf{F}_{\mathrm{BB}})_{l,:}\right\|_2^2, \qquad (32)$$

where $l = \left\lceil i\frac{N_{\mathrm{RF}}^t}{N_t}\right\rceil$. This is basically a vector approximation problem using phase rotation, and there exists a closed-form expression for nonzero elements in $\mathbf{F}_{\mathrm{RF}}$, given by

$$\arg\left\{(\mathbf{F}_{\mathrm{RF}})_{i,l}\right\} = \arg\left\{(\mathbf{F}_{\mathrm{opt}})_{i,:}(\mathbf{F}_{\mathrm{BB}})_{l,:}^H\right\},$$
$$1 \leq i \leq N_t, l = \left\lceil i\frac{N_{\mathrm{RF}}^t}{N_t}\right\rceil. \qquad (33)$$

We note that the special characteristic of $\mathbf{F}_{\mathrm{RF}}$ simplifies the analog precoder design and makes the unit modulus constraint no longer an intractable issue in the partially-connected structure.

B. Digital Baseband Precoder Design

According to (30), the precoder design at the transmit side can be rewritten as the following problem

$$
\begin{aligned}
&\underset{\mathbf{F}_{\mathrm{BB}}}{\text{minimize}} \quad \|\mathbf{F}_{\mathrm{opt}} - \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_F^2 \\
&\text{subject to} \quad \|\mathbf{F}_{\mathrm{BB}}\|_F^2 = \frac{N_{\mathrm{RF}}^t N_s}{N_t}.
\end{aligned}
\tag{34}
$$

Problem (34) is a non-convex quadratic constraint quadratic programming (QCQP) problem, which can be reformulated as a homogeneous QCQP problem:

$$
\begin{aligned}
&\underset{\mathbf{Y} \in \mathbb{H}^n}{\text{minimize}} \quad \mathrm{Tr}(\mathbf{CY}) \\
&\text{subject to} \quad
\begin{cases}
\mathrm{Tr}(\mathbf{A}_1\mathbf{Y}) = \frac{N_{\mathrm{RF}}^t N_s}{N_t} \\
\mathrm{Tr}(\mathbf{A}_2\mathbf{Y}) = 1 \\
\mathbf{Y} \succeq 0, \ \mathrm{rank}(\mathbf{Y}) = 1,
\end{cases}
\end{aligned}
\tag{35}
$$

with $\mathbb{H}^n$ being the set of $n = N_{\mathrm{RF}}^t N_s + 1$ dimension complex Hermitian matrices. In addition, $\mathbf{y} = \begin{bmatrix} \mathrm{vec}(\mathbf{F}_{\mathrm{BB}}) & t \end{bmatrix}^T$ with an auxiliary variable t, $\mathbf{Y} = \mathbf{yy}^H$, $\mathbf{f} = \mathrm{vec}(\mathbf{F}_{\mathrm{opt}})$, and

$$
\mathbf{A}_1 = \begin{bmatrix} \mathbf{I}_{n-1} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix}, \mathbf{A}_2 = \begin{bmatrix} \mathbf{0}_{n-1} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix},
$$

$$
\mathbf{C} = \begin{bmatrix} (\mathbf{I}_{N_s} \otimes \mathbf{F}_{\mathrm{RF}})^H(\mathbf{I}_{N_s} \otimes \mathbf{F}_{\mathrm{RF}}) & -(\mathbf{I}_{N_s} \otimes \mathbf{F}_{\mathrm{RF}})^H\mathbf{f} \\ -\mathbf{f}^H(\mathbf{I}_{N_s} \otimes \mathbf{F}_{\mathrm{RF}}) & \mathbf{f}^H\mathbf{f} \end{bmatrix}.
$$

The derivation and the formulation of the homogeneous QCQP problem can be found in Appendix A.

In fact, the most difficult part in problem (35) is the rank constraint, which is non-convex with respect to $\mathbf{Y}$. Thus, we first drop it to obtain a relaxed version of (35), i.e., a semidefinite relaxation (SDR) problem as follows.

$$
\begin{aligned}
&\underset{\mathbf{Y} \in \mathbb{H}^n}{\text{minimize}} \quad \mathrm{Tr}(\mathbf{CY}) \\
&\text{subject to} \quad
\begin{cases}
\mathrm{Tr}(\mathbf{A}_1\mathbf{Y}) = \frac{N_{\mathrm{RF}}^t N_s}{N_t} \\
\mathrm{Tr}(\mathbf{A}_2\mathbf{Y}) = 1 \\
\mathbf{Y} \succeq 0.
\end{cases}
\end{aligned}
\tag{36}
$$

It has been established that the SDR is tight when the number of constraints is less than three for a complex-valued homogeneous QCQP problem [51]. Consequently, problem (36) without the rank-one constraint reduces into a semidefinite programming (SDP) problem and it can be solved by standard convex optimization algorithms [52], from which we can obtain the globally optimal solution of the digital precoder design problem (34). Therefore, a step-by-step summary is provided below as the **SDR-AltMin Algorithm**.

SDR-AltMin Algorithm: Semidefinite Relaxation Based Hybrid Precoding for the Partially-connected Structure

---

Input: $\mathbf{F}_{\mathrm{opt}}$
1: Construct $\mathbf{F}_{\mathrm{RF}}^{(0)}$ with random phases and set $k = 0$;
2: **repeat**
3:     Fix $\mathbf{F}_{\mathrm{RF}}^{(k)}$, solving $\mathbf{F}_{\mathrm{BB}}^{(k)}$ using SDR (36);
4:     Fix $\mathbf{F}_{\mathrm{BB}}^{(k)}$, and update $\mathbf{F}_{\mathrm{RF}}^{(k+1)}$ by (33);
5:     $k \leftarrow k + 1$;
6: **until** a stopping criterion triggers.

C. Comparison Between Two Hybrid Precoding Structures

The main difference between two hybrid precoding structures considered in this paper is the number of phase shifters N_{PS} in use for given numbers of data streams, RF chains, and antennas.

In terms of spectral efficiency, the fully-connected structure provides more design degrees of freedom in the RF domain and thus will outperform the partially-connected one. However, when taking power consumption into consideration, it is intriguing to know which structure has better energy efficiency. Energy efficiency is defined as the ratio between spectral efficiency and total power consumption

$$
\eta = \frac{R}{P_{\mathrm{common}} + N_{\mathrm{RF}}^t P_{\mathrm{RF}} + N_t P_{\mathrm{PA}} + N_{\mathrm{PS}} P_{\mathrm{PS}}},
\tag{37}
$$

where the unit of η is bits/Hz/J and P_{common} is the common power of the transmitter. P_{RF}, P_{PS}, and P_{PA} are the power of each RF chain, phase shifter, and power amplifier, respectively. The number of phase shifters N_{PS} can be expressed as follows,

$$
N_{\mathrm{PS}} = \begin{cases} N_t N_{\mathrm{RF}}^t & \text{fully-connected} \\ N_t & \text{partially-connected} \end{cases}.
\tag{38}
$$

The numerical comparison will be provided in the Section VII.

VI. HYBRID PRECODING IN MMWAVE MIMO-OFDM SYSTEMS

In previous sections, we designed hybrid precoders for narrowband mmWave systems. On the other hand, the large available bandwidth is one of the unique characteristics of mmWave systems, and therefore the design of the hybrid precoders should be investigated when multicarrier techniques such as OFDM are utilized to overcome the multipath fading. In this section, we will extend the proposed AltMin algorithms to mmWave MIMO-OFDM systems.

In conventional MIMO-OFDM systems with sub-6 GHz carrier frequencies, digital precoding is performed in the frequency domain for each subcarrier, which can also be adopted in mmWave MIMO-OFDM systems. Futhermore, the digital precoding is followed by an inverse fast Fourier transform (IFFT) operation, which combines the signals of all the subcarriers together. However, since the analog precoding is a post-IFFT processing, the signals of all the subcarriers can only share one common analog precoder in mmWave MIMO-OFDM systems [22], [30]. Under this new restriction, the received signal of each subcarrier after the decoding process can then be expressed as

$$
\mathbf{y}[k] = \sqrt{\rho}\mathbf{W}_{\mathrm{BB}}^H[k]\mathbf{W}_{\mathrm{RF}}^H\mathbf{H}[k]\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}[k]\mathbf{s} + \mathbf{W}_{\mathrm{BB}}^H[k]\mathbf{W}_{\mathrm{RF}}^H\mathbf{n},
\tag{39}
$$

where $k \in [0, K-1]$ is the subcarrier index. $\mathbf{H}[k]$ is the frequency domain channel matrix for the kth subcarrier, which is given by [22]

$$
\mathbf{H}[k] = \gamma \sum_{i=0}^{N_{cl}-1} \sum_{l=1}^{N_{ray}} \alpha_{il}\mathbf{a}_r(\phi_{il}^r, \theta_{il}^r)\mathbf{a}_t(\phi_{il}^t, \theta_{il}^t)^H e^{-j2\pi ik/K},
\tag{40}
$$

where $\gamma = \sqrt{\frac{N_t N_r}{N_{cl} N_{ray}}}$ is the normalization factor and K is the total number of subcarriers. Then the hybrid precoder design in mmWave MIMO-OFDM systems can be formulated as [22]

$$\underset{\mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{BB}}[k]}{\text{minimize}} \quad \sum_{k=0}^{K-1} \| \mathbf{F}_{\mathrm{opt}}[k] - \mathbf{F}_{\mathrm{RF}} \mathbf{F}_{\mathrm{BB}}[k] \|_F^2 \tag{41}$$

$$\text{subject to} \quad \begin{cases} \mathbf{F}_{\mathrm{RF}} \in \mathcal{A} \\ \| \mathbf{F}_{\mathrm{RF}} \mathbf{F}_{\mathrm{BB}}[k] \|_F^2 = N_s, \end{cases}$$

where $\mathbf{F}_{\mathrm{opt}}[k]$ is the optimal digital precoder for the kth subcarrier. While it does not directly maximize the spectral efficiency, similar to the narrowband case, as shown in [13], [22], the objective is a good surrogate and it will make the problem tractable.

The alternating minimization framework can be adopted to solve problem (41). In particular, the digital precoders of all the subcarriers can be updated in a parallel fashion, since we can get rid of the summation in (41) when optimizing the digital precoder for each subcarrier. Hence, the solutions from (7), (28) and (36) still hold for problem (41). Next we will focus on the analog precoder design, which is the main difference from narrowband systems. Then the proposed AltMin algorithms can be applied to the hybrid precoding in OFDM systems.

For the MO-AltMin algorithm, based on the principles of manifold optimization, as mentioned in Section III, we first derive the Euclidean gradient of the objective function in (41) as

$$\nabla f(\mathbf{x}) = -2 \sum_{k=0}^{K-1} (\mathbf{F}_{\mathrm{BB}}^*[k] \otimes \mathbf{I}_{N_t}) \\ \times \left[\mathrm{vec}(\mathbf{F}_{\mathrm{opt}}[k]) - (\mathbf{F}_{\mathrm{BB}}^T[k] \otimes \mathbf{I}_{N_t}) \mathbf{x} \right]. \tag{42}$$

After calculating this Euclidean gradient, we can still use the projection (13) and the retraction (15) to obtain the Riemannian gradient and the mapped gradient vector on the manifold, which are the key elements in Algorithm 1.

For the PE-AltMin algorithm, after we adopt a similar upper bound and manipulations in Section IV, the problem formulation for the analog precoder design is given as

$$\underset{\mathbf{F}_{\mathrm{RF}}}{\text{minimize}} \quad \sum_{k=0}^{K-1} \left\| \mathbf{F}_{\mathrm{opt}}[k] \mathbf{F}_{\mathrm{DD}}^H[k] - \mathbf{F}_{\mathrm{RF}} \right\|_F^2 \tag{43}$$

$$\text{subject to} \quad |(\mathbf{F}_{\mathrm{RF}})_{i,j}| = 1, \forall i, j,$$

which has a closed-form solution

$$\arg (\mathbf{F}_{\mathrm{RF}}) = \arg \left(\sum_{k=0}^{K-1} \mathbf{F}_{\mathrm{opt}}[k] \mathbf{F}_{\mathrm{DD}}^H[k] \right). \tag{44}$$

By substituting Step (5) in the PE-AltMin algorithm, this solution enables the extension of the PE-AltMin algorithm to OFDM systems.

Similar to the previous two AltMin Algorithms, the SDR-AltMin algorithm can also be realized in OFDM systems. We

can get the closed-form solution for the analog precoder as

$$\arg \{(\mathbf{F}_{\mathrm{RF}})_{i,l}\} = \arg \left\{ \sum_{k=0}^{K-1} (\mathbf{F}_{\mathrm{opt}}[k])_{i,:} (\mathbf{F}_{\mathrm{BB}}[k])_{l,:}^H \right\},$$

$$1 \le i \le N_t, l = \left\lceil i \frac{N_{\mathrm{RF}}^t}{N_t} \right\rceil, \tag{45}$$

which can be added in Step 4 in the SDR-AltMin algorithm.

The aforementioned solutions demonstrate that the proposed AltMin algorithms can be directly extended to mmWave MIMO-OFDM systems, and the performance of the proposed AltMin algorithms will be evaluated in the next section.

VII. SIMULATION RESULTS

In this section, we will numerically evaluate the performance of our proposed algorithms. Data streams are sent from a transmitter with $N_t = 144$ to a receiver with $N_r = 36$ antennas, while both are equipped with USPA. The channel parameters are set as $N_{cl} = 5$ clusters, $N_{ray} = 10$ rays and the average power of each cluster is $\sigma_{\alpha,i}^2 = 1$. The azimuth and elevation AoDs and AoAs follow the Laplacian distribution with uniformly distributed mean angles over $[0, 2\pi)$ and angular spread of 10 degrees. The antenna elements in the USPA are separated by a half wavelength distance and all simulation results are averaged over 1000 channel realizations. For all the proposed AltMin algorithms, the initial phases of the analog precoder $\mathbf{F}_{\mathrm{RF}}$ follow a uniform distribution over $[0, 2\pi)$.

A. Spectral Efficiency Evaluation

Firstly, we investigate the spectral efficiency achieved by different algorithms when the number of RF chains is equal to that of the data streams, i.e., $N_{\mathrm{RF}}^t = N_{\mathrm{RF}}^r = N_s$. This is the worst case since the number of RF chains cannot be smaller under the assumptions in Section II-A. In this case, as shown in Fig. 3, for the fully-connected structure, the existing OMP algorithm [13] achieve significantly lower spectral efficiency than the optimal digital precoder. On the contrary, our proposed alternating minimization algorithm[4], i.e., the MO-AltMin algorithm, achieves near-optimal performance over the whole SNR range in consideration. This means that the proposed algorithm can more accurately approximate the optimal digital precoder than existing algorithms, even though the RF chains are limited. In this scenario, the partially-connected structure with the proposed SDR-AltMin algorithm provides substantial performance gains over analog beamforming, especially at high SNRs. For comparison[5], the SIC-Based method proposed in [31] is adopted as a benchmark for the partially-connected structure. Fig. 3 shows

---

[4]When optimizing the analog precoder, Algorithm 1 converges within 20 iterations for almost all the channel realizations.

[5]Since the SIC-Based method in [31] can only design hybrid precoders at the transmitter, we assume that the optimal digital decoder is adopted at the receiver. Furthermore, the same setting is also employed for the SDR-AltMin algorithm for fair comparison in Fig. 3. Nevertheless, in the remainder of this section, both hybrid precoder and decoder are assumed at the transmitter and receiver for the partially-connected structure.
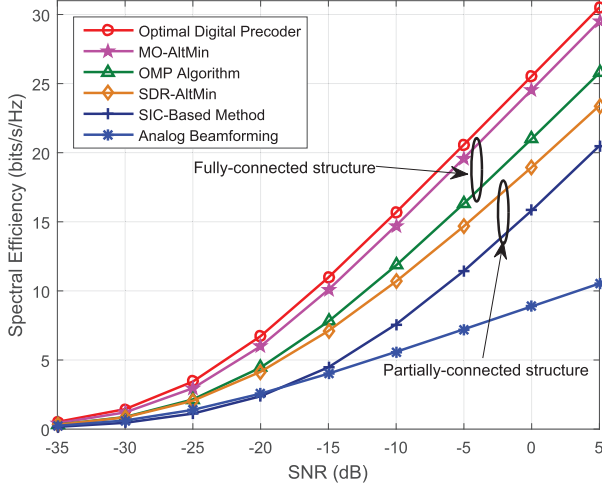
Fig. 3. Spectral efficiency achieved by different precoding algorithms when $N_{\mathrm{RF}}^t = N_{\mathrm{RF}}^r = N_s = 3$.
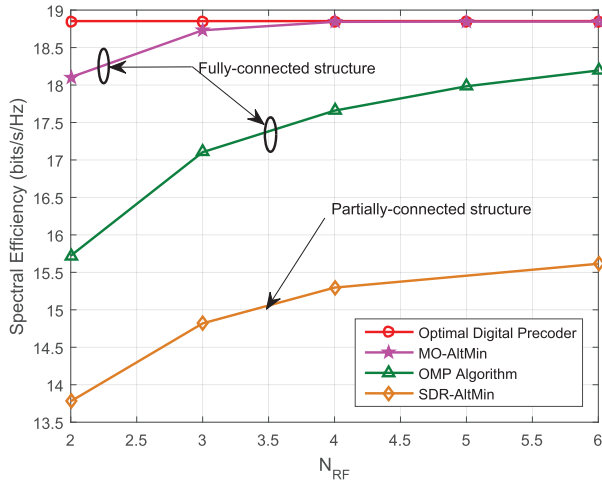

Fig. 4. Spectral efficiency achieved by different precoding algorithms given $N_s = 2$, $N_{\mathrm{RF}}^t = N_{\mathrm{RF}}^r = N_{\mathrm{RF}}$ and SNR $= 0$ dB.

Fig. 5. Energy efficiency of the fully-connected and partially-connected structures when $N_s = 2$, $N_{\mathrm{RF}}^t = N_{\mathrm{RF}}^r = N_{\mathrm{RF}}$ and SNR $= 0$ dB.

partially-connected structure, using less phase shifters, does entail some non-negligible performance loss when compared with the fully-connected structure.

B. Energy Efficiency Evaluation

In this part, we will compare the performance of the two hybrid precoding structures in terms of energy efficiency, as defined in (37). The simulation parameters are set as follows: $P_{\mathrm{common}} = 10$ W, $P_{\mathrm{RF}} = 100$ mW, $P_{\mathrm{PS}} = 10$ mW and $P_{\mathrm{PA}} = 100$ mW [12]. The simulation results are shown in Fig. 5, which shows substantially different behaviors for the two structures. Since the number of phase shifters scales linearly with N_{RF}^t and N_t in the fully-connected structure, the power consumption will increase substantially when increasing N_{RF}^t. As shown in Fig. 4, however, the spectral efficiency achieved by the proposed MO-AltMin algorithm is sufficiently close or exactly equal to the optimal digital one, and will not increase further as N_{RF} increases. Based on these two facts, the power consumption grows much faster than the spectral efficiency, which gives rise to the dramatic decrease of the energy efficiency.

For the partially-connected structure, as the number of phase shifters is independent of N_{RF}^t, the dominant part of total power consumption remains almost unchanged over the investigated range of RF chain numbers. Meanwhile, the spectral efficiency will gradually approach the optimal digital precoder when increasing N_{RF}^t. The improvement of the spectral efficiency and the almost unchanged power consumption together account for the rise in the energy efficiency when N_{RF}^t goes up in the partially-connected structure.

More importantly, Fig. 5 shows that there is an intersection point, i.e., when $N_{\mathrm{RF}} = 5$, of the energy efficiency for two hybrid precoding structures. In particular, the fully-connected structure enjoys a higher energy efficiency with a small number of RF chains, while the partially-connected one is more energy efficient when a relatively large number of RF chains are implemented at transceivers. This phenomenon will offer valuable insights for the RF chain implementation in hybrid precoding. As shown in Fig. 4, the fully-connected structure can approach the performance of the optimal digital precoder

that the SDR-AltMin algorithm outperforms the benchmark. This is mainly because the SDR-AltMin algorithm takes full use of the digital precoder, while the SIC-Based method only uses the digital precoder to allocate power to the data streams under the extra diagonal constraint.

It has been shown in [26] that when $N_{\mathrm{RF}}^t \geq 2N_s$ and $N_{\mathrm{RF}}^r \geq 2N_s$, there exists a closed-form solution to the design problem of the fully-connected hybrid precoding, which leads to the same spectral efficiency provided by the optimal digital precoder. Although the hybrid precoder design in this paper aims at cases of $N_s \leq N_{\mathrm{RF}}^t < 2N_s$, it is interesting to examine if our proposed algorithm can achieve the comparable performance as the case considered in [26]. Fig. 4 compares the performance of different precoding schemes for different N_{RF}. We see that the proposed MO-AltMin algorithm for the fully-connected structure starts to coincide with the optimal digital precoding when $N_{\mathrm{RF}}^t = N_{\mathrm{RF}}^r \geq 4$. This result demonstrates that when $N_{\mathrm{RF}}^t \geq 2N_s$ and $N_{\mathrm{RF}}^r \geq 2N_s$, our proposed algorithm can actually achieve the optimal spectral efficiency, which, however, cannot be achieved by the OMP algorithm. Furthermore, the comparison between two hybrid precoding structures shows that the
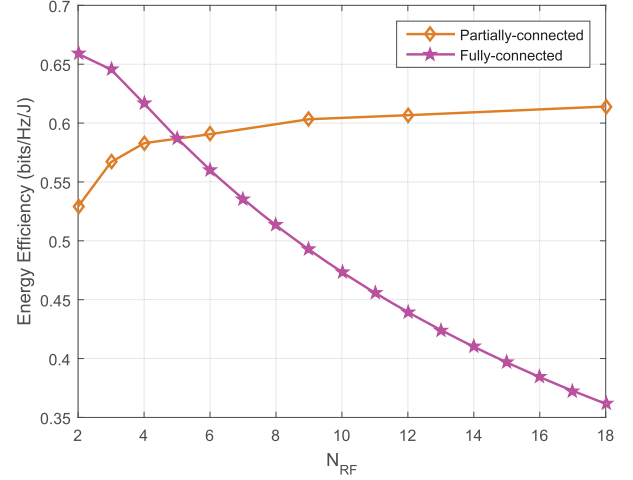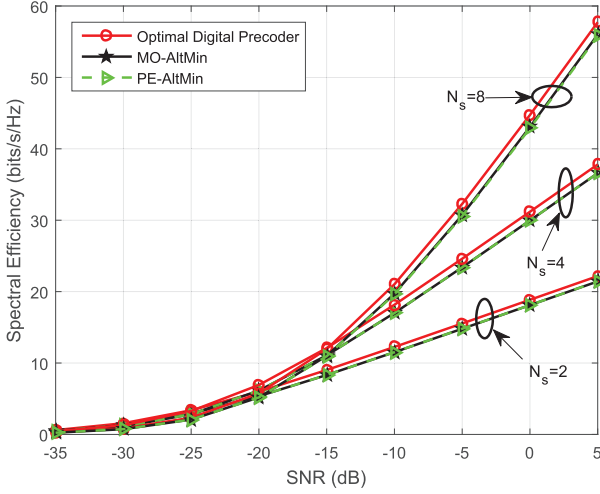
Fig. 6. Spectral efficiency achieved by the MO-AltMin and PE-AltMin algorithms given $N_{\mathrm{RF}}^t = N_{\mathrm{RF}}^r = N_{\mathrm{RF}} = N_s$.

Fig. 7. Spectral efficiency achieved by different precoding algorithms given $N_s = 6$, $N_{\mathrm{RF}}^t = N_{\mathrm{RF}}^r = N_{\mathrm{RF}}$ and SNR = 0 dB.

Fig. 8. Spectral efficiency achieved by different precoding algorithms in mmWave MIMO-OFDM systems given $N_{\mathrm{RF}}^t = N_{\mathrm{RF}}^r = N_{\mathrm{RF}}$.

when the number of RF chains is slightly larger than that of the data streams. Therefore, there is no need to implement more RF chains considering the energy efficiency. On the other hand, with a low-complexity hardware implementation, it is beneficial for the partially-connected structure to leverage the larger size of RF chains to improve both spectral and energy efficiency.

C. Low-Complexity Design for the Fully Connected Structure

As mentioned in Section IV-B, there are two intermediate steps taken for developing the PE-AltMin algorithm, i.e., constructing the digital precoder with orthogonal columns and replacing the objective function with an upper bound. Firstly, we test the influence of the additional constraint on the digital precoder, for which we eliminate the impact of the second step by setting the numbers of RF chains and data streams to be equal, i.e., to make the bound tight.

Fig. 6 plots the spectral efficiency achieved by both the MO-AltMin and PE-AltMin algorithms when there are 2, 4 and 8 data streams transmitted. We see that the curves of the low-complexity PE-AltMin algorithm nearly coincide with those of the MO-AltMin algorithm. This phenomenon implies that the orthogonal column structure of the digital precoder has negligible impact on spectral efficiency, which justifies the rationality of the digital precoder design in Section IV-B. Fig. 6 also indicates that we can achieve the performance of the high complexity MO-AltMin algorithm by adopting the low-complexity PE-AltMin algorithm when $N_{\mathrm{RF}} = N_s$. Under this system setting, the PE-AltMin algorithm serves as an excellent candidate for the hybrid precoder design, achieving both good performance and low complexity. On the contrary, the OMP algorithm works poorly when $N_{\mathrm{RF}} = N_s$.

Next we investigate the impact of the number of RF chains. Fig. 7 compares different algorithms assuming 6 data streams are transmitted. Since the optimal solution in the $N_{\mathrm{RF}}^t \geq 2N_s$ region has been fully developed in [26], here we focus on the remaining region, i.e., $N_{\mathrm{RF}}^t = N_{\mathrm{RF}}^r = N_{\mathrm{RF}} \in [6, 11]$. From Fig. 7, we find that the PE-AltMin algorithm has a small gap compared with the MO-AltMin algorithm. That is because the
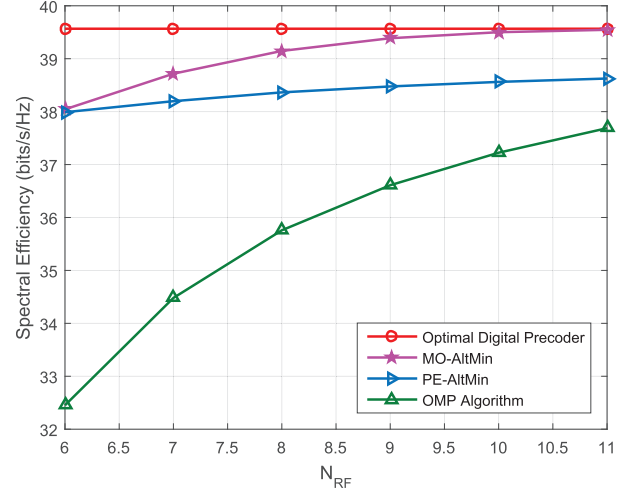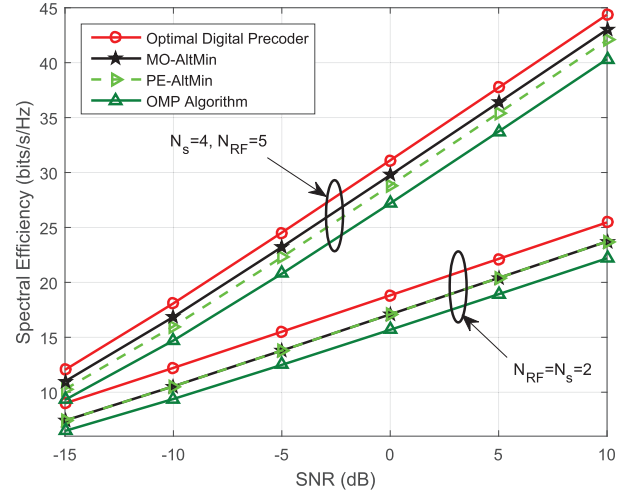
low-complexity algorithm tries to minimize an upper bound instead of the original objective function. As explained in Section IV-B, the upper bound is tight when $N_{\mathrm{RF}} = N_s$ and gets looser when N_{RF} increases, which determines the gap between the MO-AltMin and PE-AltMin algorithms. However, the spectral efficiency provided by the PE-AltMin algorithm is far higher than the existing OMP algorithm, especially when the RF chain number is small. With the proposed AltMin algorithms, we see that the performance of the fully-connected structure approaches that of the fully digital precoder when the number of RF chains is comparable with the number of data streams, which cannot be revealed from the OMP algorithm.

D. Hybrid Precoding in MmWave MIMO-OFDM Systems

In this part, we will show the performance of the proposed AltMin algorithms when applied to mmWave MIMO-OFDM systems. We assume that the number of subcarriers is $K = 128$.

Fig. 8 plots the spectral efficiency achieved by the MO-AltMin and PE-AltMin algorithms compared to the OMP-based

method proposed in [22]. It turns out that the MO-AltMin algorithm always has the highest spectral efficiency under different system parameters. With a large number of RF chains, similar to what we have observed in Fig. 7, the MO-AltMin algorithm will quickly get close to the performance of the optimal digital precoder. Interestingly, we discover that the low-complexity PE-AltMin algorithm can achieve almost the same spectral efficiency as that of the MO-AltMin algorithm when the numbers of RF chains and data streams are equal. This phenomenon is the same as that in narrowband systems, and it demonstrates that the extra orthogonality constraint on the digital precoder also has negligible impact on spectral efficiency in mmWave OFDM systems. It is also found in Fig. 8 that the PE-AltMin algorithm outperforms the OMP-based method in mmWave OFDM systems. This indicates that the PE-AltMin algorithm can serve as an outstanding candidate for the low-complexity hybrid precoding, both in narrowband and broadband OFDM systems, when the transceivers only have limited RF chains available.

VIII. Conclusions

Built on the principle of alternating minimization, in this paper, we proposed an innovative design methodology for hybrid precoding in mmWave MIMO systems. Effective algorithms were proposed for the fully-connected and partially-connected hybrid precoding structures, and simulation results helped reveal the following valuable design insights:

- The hybrid precoders with the fully-connected structure can approach the performance of the fully digital precoder when the number of RF chains is slightly larger than the number of data streams. Considering the increasing cost and power consumption, there is no need to further increase the number of RF chains.
- For the partially-connected structure, in terms of spectral efficiency, hybrid precoders provide substantial gains over analog beamforming. Furthermore, it is profitable to implement a relatively large number of RF chains, in order to enhance both spectral and energy efficiency.

Finally, our results have clearly demonstrated the effectiveness of alternating minimization in designing hybrid precoders in mmWave MIMO systems. It will be interesting to extend the alternating minimization techniques to other hybrid precoder design problems, as well as to consider the hybrid precoder design combined with channel training and feedback. Also a finer convergence analysis and optimality characterization of the proposed algorithms will require further investigation.

Appendix A
Formulation of the Homogeneous QCQP Problem

The original problem is a non-homogeneous QCQP problem

$$\underset{\mathbf{F}_{\mathrm{BB}}}{\text{minimize}} \quad \|\mathbf{F}_{\mathrm{opt}} - \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_F^2$$
$$\text{subject to} \quad \|\mathbf{F}_{\mathrm{BB}}\|_F^2 = \frac{N_{\mathrm{RF}}^t N_s}{N_t}. \tag{46}$$

According to the vectorization property, the objective function in (46) can be rewritten as

$$\|\mathbf{F}_{\mathrm{opt}} - \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_F^2 = \|\mathrm{vec}(\mathbf{F}_{\mathrm{opt}} - \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}})\|_2^2$$
$$= \|\mathrm{vec}(\mathbf{F}_{\mathrm{opt}}) - \mathrm{vec}(\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}})\|_2^2$$
$$= \|\mathrm{vec}(\mathbf{F}_{\mathrm{opt}}) - (\mathbf{I}_{N_s} \otimes \mathbf{F}_{\mathrm{RF}})\mathrm{vec}(\mathbf{F}_{\mathrm{BB}})\|_2^2. \tag{47}$$

To simplify the notation, denote $\mathbf{f} = \mathrm{vec}(\mathbf{F}_{\mathrm{opt}})$, $\mathbf{b} = \mathrm{vec}(\mathbf{F}_{\mathrm{BB}})$ and $\mathbf{E} = \mathbf{I}_{N_s} \otimes \mathbf{F}_{\mathrm{RF}}$. In order to apply SDR, we introduce an auxiliary variable t to homogenize the original problem as

$$\underset{\mathbf{b}}{\text{minimize}} \quad \|t\mathbf{f} - \mathbf{Eb}\|_2^2$$
$$\text{subject to} \quad \begin{cases} \|\mathbf{b}\|_2^2 = \frac{N_{\mathrm{RF}}^t N_s}{N_t} \\ t^2 = 1. \end{cases} \tag{48}$$

The objective function in (48) can be further rewritten as

$$\|t\mathbf{f} - \mathbf{Eb}\|_2^2 = \begin{bmatrix} \mathbf{b}^H & t \end{bmatrix} \begin{bmatrix} \mathbf{E}^H\mathbf{E} & -\mathbf{E}^H\mathbf{f} \\ -\mathbf{f}^H\mathbf{E} & \mathbf{f}^H\mathbf{f} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ t \end{bmatrix}. \tag{49}$$

Furthermore, the two constraints in (48) can also be manipulated as

$$\|\mathbf{b}\|_2^2 = \begin{bmatrix} \mathbf{b}^H & t \end{bmatrix} \begin{bmatrix} \mathbf{I}_{N_{\mathrm{RF}}^t N_s} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ t \end{bmatrix} = \frac{N_{\mathrm{RF}}^t N_s}{N_t}, \tag{50}$$

$$t^2 = \begin{bmatrix} \mathbf{b}^H & t \end{bmatrix} \begin{bmatrix} \mathbf{0}_{N_{\mathrm{RF}}^t N_s} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ t \end{bmatrix} = 1. \tag{51}$$

Consequently, if we define

$$\mathbf{y} = \begin{bmatrix} \mathbf{b} \\ t \end{bmatrix}, \mathbf{Y} = \mathbf{y}\mathbf{y}^H, \mathbf{C} = \begin{bmatrix} \mathbf{E}^H\mathbf{E} & -\mathbf{E}^H\mathbf{f} \\ -\mathbf{f}^H\mathbf{E} & \mathbf{f}^H\mathbf{f} \end{bmatrix},$$
$$\mathbf{A}_1 = \begin{bmatrix} \mathbf{I}_{N_{\mathrm{RF}}^t N_s} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix}, \mathbf{A}_2 = \begin{bmatrix} \mathbf{0}_{N_{\mathrm{RF}}^t N_s} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix},$$

the original problem (46) can be reformulated as

$$\underset{\mathbf{Y} \in \mathbb{H}^n}{\text{minimize}} \quad \mathrm{Tr}(\mathbf{CY})$$
$$\text{subject to} \quad \begin{cases} \mathrm{Tr}(\mathbf{A}_1\mathbf{Y}) = \frac{N_{\mathrm{RF}}^t N_s}{N_t} \\ \mathrm{Tr}(\mathbf{A}_2\mathbf{Y}) = 1 \\ \mathbf{Y} \succeq 0, \ \mathrm{rank}(\mathbf{Y}) = 1. \end{cases} \tag{52}$$

References

[1] X. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, "Hybrid precoding design in millimeter wave MIMO systems: An alternating minimization approach," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, San Diego, CA, USA, Dec. 2015, pp. 1–6.

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

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

[4] C. Li, J. Zhang, and K. B. Letaief, "Throughput and energy efficiency analysis of small cell networks with multi-antenna base stations," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2502–2517, May 2014.

[5] G. Bartoli *et al.*, "Beamforming for small cell deployment in LTE-advanced and beyond," *IEEE Wireless Commun.*, vol. 21, no. 2, pp. 50–56, Apr. 2014.

[6] K. Doppler, M. Rinne, C. Wijting, C. B. Ribeiro, and K. Hugl, "Device-to-device communication as an underlay to LTE-advanced networks," *IEEE Commun. Mag.*, vol. 47, no. 12, pp. 42–49, Dec. 2009.

[7] Y. Shi, J. Zhang, K. B. Letaief, B. Bai, and W. Chen, "Large-scale convex optimization for ultra-dense Cloud-RAN," *IEEE Wireless Commun.*, vol. 22, no. 3, pp. 84–91, Jun. 2015.

[8] Y. Shi, J. Zhang, and K. B. Letaief, "Group sparse beamforming for green Cloud-RAN," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2809–2823, May 2014.

[9] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.

[10] E. Torkildson, U. Madhow, and M. Rodwell, "Indoor millimeter wave MIMO: Feasibility and performance," *IEEE Trans. Wireless Commun.*, vol. 10, no. 12, pp. 4150–4160, Dec. 2011.

[11] M. R. Akdeniz *et al.*, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[12] T. S. Rappaport, R. W. Heath, R. C. Daniels, and J. N. Murdock, *Millimeter Wave Wireless Communications*. New York, NY, USA: Pearson Education, 2014.

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

[14] P. Jain, P. Netrapalli, and S. Sanghavi, "Low-rank matrix completion using alternating minimization," in *Proc. 45th ACM Symp. Theory Comput. (STOC)*, Palo Alto, CA, USA, Jun. 2013, pp. 665–674.

[15] P. Drineas and M. W. Mahoney, "On the Nyström method for approximating a gram matrix for improved Kernel-based learning," *J. Mach. Learn. Res.*, vol. 6, pp. 2153–2175, Dec. 2005.

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

[17] S. Sun, T. S. Rappaport, R. W. Heath, A. Nix, and S. Rangan, "MIMO for millimeter-wave wireless communications: Beamforming, spatial multiplexing, or both?" *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 110–121, Dec. 2014.

[18] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[19] P. Wang, Y. Li, L. Song, and B. Vucetic, "Multi-gigabit millimeter wave wireless communications for 5G: From fixed access to cellular networks," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 168–178, Jan. 2015.

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

[21] Y. Lee, C.-H. Wang, and Y.-H. Huang, "A hybrid RF/baseband precoding processor based on parallel-index-selection matrix-inversion-bypass simultaneous orthogonal matching pursuit for millimeter wave MIMO systems," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 305–317, Jan. 2015.

[22] J. Lee and Y. Lee, "AF relaying for millimeter wave communication systems with hybrid RF/baseband MIMO processing," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Sydney, NSW, Australia, Jun. 2014, pp. 5838–5842.

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

[24] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3814–3827, Jul. 2013.

[25] C. Rusu, R. Mendez-Rial, N. Gonzalez-Prelcicy, and R. W. Heath, "Low complexity hybrid sparse precoding and combining in millimeter wave MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., Jun. 2015, pp. 1340–1345.

[26] E. Zhang and C. Huang, "On achieving optimal rate of digital precoder by RF-baseband codesign for MIMO systems," in *Proc. 80th IEEE Veh. Technol. Conf. (VTC-Fall)*, Vancouver, BC, Canada, Sep. 2014, pp. 1–5.

[27] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 653–656, Dec. 2014.

[28] G. Wang and G. Ascheid, "Joint pre/post-processing design for large millimeter wave hybrid spatial processing systems," in *Proc. 20th Eur. Wireless Conf.*, Barcelona, Spain, May 2014, pp. 1–6.

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

[30] C. Kim, T. Kim, and J.-Y. Seol, "Multi-beam transmission diversity with hybrid beamforming for MIMO-OFDM systems," in *Proc. IEEE Global Commun. Conf. Workshops (GLOBECOM)*, Atlanta, GA, USA, Dec. 2013, pp. 61–65.

[31] L. Dai, X. Gao, J. Quan, S. Han, and C.-L. I, "Near-optimal hybrid analog and digital precoding for downlink mmwave massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., Jun. 2015, pp. 1334–1339.

[32] 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.

[33] J. Zhang, X. Huang, V. Dyadyuk, and Y. Guo, "Massive hybrid antenna array for millimeter-wave cellular communications," *IEEE Wireless Commun.*, vol. 22, no. 1, pp. 79–87, Feb. 2015.

[34] A. Abbaspour-Tamijani and K. Sarabandi, "An affordable millimeter-wave beam-steerable antenna using interleaved planar subarrays," *IEEE Trans. Antennas Propag.*, vol. 51, no. 9, pp. 2193–2202, Nov. 2003.

[35] D. J. Love and R. W. Heath, "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2967–2976, Aug. 2005.

[36] P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Dec. 2013, pp. 2796–2804.

[37] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM J. Imag. Sci.*, vol. 1, no. 3, pp. 248–272, Aug. 2008.

[38] T. F. Chan and C.-K. Wong, "Convergence of the alternating minimization algorithm for blind deconvolution," *Linear Algebra Appl.*, vol. 316, no. 1, pp. 259–285, Apr. 2000.

[39] H. Kim and H. Park, "Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 2, pp. 713–730, Jul. 2008.

[40] X. Zhang, A. F. Molisch, and S.-Y. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4091–4103, Nov. 2005.

[41] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ, USA: Princeton Univ. Press, 2009.

[42] Y. Ma and Y. Fu, *Manifold Learning Theory and Applications*. Boca Raton, FL, USA: CRC Press, 2012.

[43] J. Lee, *Introduction to Smooth Manifolds*. New York, NY, USA: Springer, 2012.

[44] Y. Shi, J. Zhang, and K. B. Letaief, "Low-rank matrix completion via Riemannian pursuit for topological interference management," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Hong Kong, Jun. 2015, pp. 1831–1835.

[45] A. Hjorungnes, *Complex-Valued Matrix Derivatives*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[46] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, "Manopt, a Matlab toolbox for optimization on manifolds," *J. Mach. Learn. Res.*, vol. 15, pp. 1455–1459, Jan. 2014.

[47] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA, USA: Athena Scientific, 1999.

[48] A. Patrascu and I. Necoara, "Efficient random coordinate descent algorithms for large-scale structured nonconvex optimization," *J. Global Optim.*, vol. 61, no. 1, pp. 19–46, Feb. 2015.

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

[50] J. C. Gower and G. B. Dijksterhuis, *Procrustes Problems*. London, U.K.: Oxford Univ. Press, 2004.

[51] Z.-Q. Luo, W.-K. Ma, A.-C. 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.

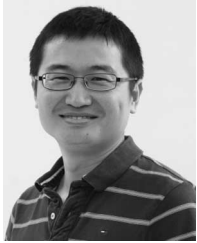[52] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
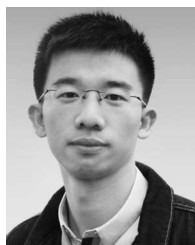
Xianghao Yu (S'15) received the B.Eng. degree in information engineering from Southeast University (SEU), Nanjing, China, in 2014. He is currently pursuing the Ph.D. degree in electronic and computer engineering at the Hong Kong University of Science and Technology (HKUST), Hong Kong, China. His research interests include millimeter wave communications, MIMO systems, mathematical optimization, and stochastic geometry.

Juei-Chin Shen (S'10–M'14) received the Ph.D. degree in communication engineering from the University of Manchester, Manchester, U.K., in 2013. He was with the Hong Kong University of Science and Technology (HKUST) from 2013 to 2015, as a Research Associate. Since 2015, he has been with Mediatek Inc. as a Senior Engineer, working on the research and standardization of future wireless technologies. His research interests include compressed sensing, massive MIMO systems, and millimeter-wave communications. He was a corecipient of the IEEE PIMRC 2014 Best Paper Award.

Jun Zhang (S'06–M'10–SM'15) received the B.Eng. degree in electronic engineering from the University of Science and Technology of China, Hefei, China, in 2004, the M.Phil. degree in information engineering from the Chinese University of Hong Kong, Hong Kong, in 2006, and the Ph.D. degree in electrical and computer engineering from the University of Texas at Austin, Austin, TX, USA, in 2009. He is currently a Research Assistant Professor with the Department of Electronic and Computer Engineering, the Hong Kong University of Science and Technology (HKUST). He has coauthored the book *Fundamentals of LTE* (Prentice-Hall, 2010). His research interests include wireless communications and networking, green communications, and signal processing. He is an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and served as a MAC track Co-Chair for the IEEE WCNC 2011. He was the recipient of the 2014 Best Paper Award for the *EURASIP Journal on Advances in Signal Processing*, and the PIMRC 2014 Best Paper Award.

Khaled B. Letaief (S'85–M'86–SM'97–F'03) received the B.S. degree *with distinction* in electrical engineering (1984) from Purdue University, West Lafayette, IN, USA. He has also received the M.S. and Ph.D. degrees in electrical engineering from Purdue University in 1986 and 1990, respectively.

From 1990 to 1993, he was a faculty member at the University of Melbourne, Australia. He has been with the Hong Kong University of Science & Technology. While at HKUST, he has held numerous administrative positions, including the Head of the Department of Electronic and Computer Engineering, Director of the Center for Wireless IC Design, Director of Huawei Innovation Laboratory, and Director of the Hong Kong Telecom Institute of Information Technology. He has also served as Chair Professor and Dean of HKUST School of Engineering. Under his leadership, the School of Engineering has dazzled in international rankings (*ranked #14 in the world in 2015* according to QS World University Rankings).

From September 2015, he joined Hamad Bin Khalifa University (HBKU) as Provost to help establish a research-intensive university in Qatar in partnership with strategic partners that include Northwestern University, Carnegie Mellon University, Cornell, and Texas A&M.

Dr. Letaief is a world-renowned leader in wireless communications and networks. In these areas, he has over 500 journal and conference papers and given invited keynote talks as well as courses all over the world. He has made 6 major contributions to IEEE Standards along with 13 patents including 11 U.S. patents.

He served as consultants for different organizations and is the founding Editor-in-Chief of the prestigious IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He has served on the editorial board of other prestigious journals including the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS – *Wireless Series* (as Editor-in-Chief). He has been involved in organizing a number of major international conferences

Professor Letaief has been a dedicated educator committed to excellence in teaching and scholarship. He received the *Mangoon Teaching Award* from Purdue University in 1990; HKUST Engineering Teaching Excellence Award (4 times); and the Michael Gale Medal for Distinguished Teaching (*Highest university-wide teaching award* at HKUST).

He is also the recipient of many other distinguished awards including 2007 *IEEE Joseph LoCicero* Publications Exemplary Award; 2009 *IEEE Marconi Prize Award in Wireless Communications*; 2010 Purdue University Outstanding Electrical and Computer Engineer Award; 2011 IEEE Harold Sobol Award; 2011 IEEE Wireless Communications Technical Committee Recognition Award; and 11 IEEE Best Paper Awards.

He had the privilege to serve IEEE in many leadership positions including IEEE ComSoc Vice-President, IEEE ComSoc Director of Journals, and member of IEEE Publications Services and Products Board, IEEE ComSoc Board of Governors, IEEE TAB Periodicals Committee, and IEEE Fellow Committee.

Dr. Letaief is a Fellow of IEEE and a Fellow of HKIE. He is also recognized by Thomson Reuters as an *ISI Highly Cited Researcher*.