Channel-Statistics-Based Hybrid Precoding for Millimeter-Wave MIMO Systems With Dynamic Subarrays

Juening Jin, Chengshan Xiao, *Fellow, IEEE*, Wen Chen, *Senior Member, IEEE*, and Yongpeng Wu, *Senior Member, IEEE*

Abstract—This paper investigates the hybrid precoding design for millimeter wave (mmWave) multiple-input-multiple-output (MIMO) systems with finite-alphabet inputs. The mmWave MIMO system employs partially-connected hybrid precoding architecture with dynamic subarrays, where each radio frequency (RF) chain is connected to a dynamic subset of antennas. We consider the design of analog and digital precoders utilizing statistical and/or mixed channel state information (CSI), which involve solving an extremely difficult problem in theory: First, designing the optimal partition of antennas over RF chains is a combinatorial optimization problem, whose optimal solution requires an exhaustive search over all antenna partitioning solutions; Second, the average mutual information under mmWave MIMO channels lacks closed-form expression and involves prohibitive computational burden; and Third, the hybrid precoding problem with given partition of antennas is nonconvex with respect to the analog and digital precoders. To address these issues, this paper first presents a simple criterion and the corresponding low complexity algorithm to design the optimal partition of antennas using statistical CSI. Then, it derives the lower bound and its approximation for the average mutual information, in which the computational complexity is greatly reduced compared to calculating the average mutual information directly. In addition, it also shows that the lower bound with a constant shift offers a very accurate approximation to the average mutual information. This paper further proposes utilizing the lower bound approximation as a low-complexity and accurate alternative for developing a manifold-based gradient ascent algorithm to find near-optimal analog and digital precoders. Several numerical results are provided to show that our proposed algorithm outperforms the existing hybrid precoding algorithms.

Index Terms—Hybrid precoding, finite-alphabet inputs, matrix factorization, nonconvex optimization.

I. INTRODUCTION

MASSIVE multiple-input multiple-output (MIMO) systems operating in the Millimeter wave (mmWave) band is a key technique candidate for future generation cellular systems to address the wireless spectrum crunch. It makes use of the frequency band from 30 GHz to 300 GHz, which provides a much wider bandwidth than current cellular systems operating in microwave bands. In addition, a short wavelength of radio signals in the mmWave band enables very large antenna arrays to be equipped at the transceivers, and this can provide significant increase of the spectral efficiency.

For mmWave MIMO systems, hybrid analog and digital precoding architectures have been proposed to achieve high spectral efficiency with low cost and power consumption. Extensive work has been devoted to designing hybrid precoding algorithms under perfect channel state information (CSI) and different constraints [1]–[8]. However, it is difficult to obtain the perfect CSI in mmWave MIMO systems. The reason is that the channel matrix measured at the baseband cannot be obtained directly because it is intertwined with the choice of analog precoders. Furthermore, conventional MIMO channel estimation is incapable of utilizing array gain in mmWave systems, and it leads to low signal-to-noise ratio (SNR). Therefore, the conventional channel estimation requires long training sequences to estimate mmWave MIMO channels, which is impractical due to fast variation of mmWave MIMO channels.

To address the challenge of training overhead, [9] proposed a hybrid precoding algorithm for single-user MIMO systems with partial knowledge of the CSI. For the multi-user MIMO scenario, [10] devised a mix-CSI-based hybrid precoding structure, where the analog precoding design is based on the slow-varying channel statistics, and the digital precoding design is based on the instantaneous CSI. Then the dimension of the effective channel matrix (instantaneous CSI) is greatly reduced. However, the work in [9] and [10]

considered only the fully-connected hybrid architecture, which requires much more phase shifters compared to the partially-connected structure [11]. In the partially-connected structure, the antenna array is partitioned into a number of smaller disjoint subarrays, each of which is driven by a single radio frequency (RF) chain [12]. This structure is an extension of classic antenna selection methods, which allocate each RF chain to an antenna element [13]. Gao *et al.* [14] developed a successive interference cancellation based hybrid precoding for partially-connected structure with fixed subset of antennas. The partially-connected structure with dynamic subset of antennas is considered in [15], and a low complexity greedy algorithm is also proposed to design the best partitioning/grouping of antennas over the RF chains.

Furthermore, most existing works on hybrid precoding assume Gaussian inputs, which are rarely realized in practice. It is well known that practical systems utilize finite-alphabet inputs, such as phase-shift keying (PSK) or quadrature amplitude modulation (QAM). Precoding designs under Gaussian inputs have been shown to be quite suboptimal for practical systems with finite-alphabet inputs [16]–[27]. Recently, the authors in [28] presented a Broyden-Fletcher-Goldfarb-Shanno based hybrid precoding algorithm for mmWave MIMO systems with finite-alphabet inputs. The proposed algorithm utilizes both gradient and Hessian information, and simulation results showed that it outperforms existing hybrid precoding algorithms including [3], [5], [6], [8].

A. Contributions

In this paper, we investigate the hybrid precoding design for mmWave MIMO systems with finite-alphabet inputs under the following assumptions: 1) the system employs partially-connected hybrid precoding structure with dynamic subset of antennas; 2) the partition of antennas and analog precoder are designed based on statistical CSI, and the digital precoder is designed based on either statistical CSI or instantaneous CSI. We consider the statistical-CSI-based scenario and the mixed-CSI-based scenario, and the corresponding hybrid precoding problems under two scenarios have the same mathematical form. Then we propose a manifold-based gradient ascent algorithm to solve the hybrid precoding problem. The contributions of this paper are summarized as follows:

- We present a simple criterion to design the best partition of antennas using statistical CSI. The corresponding dynamic subarray design is a (nonconvex) combinatorial optimization problem, and we propose a low complexity algorithm to solve this problem.
- We derive a lower bound of the average mutual information for mmWave MIMO channels. The lower bound plus a constant shift serves as a very accurate approximation to the average mutual information, and its complexity is much lower than the original average mutual information. To further reduce the complexity, we also derive an accurate approximation of the proposed lower bound.
- We propose a manifold-based gradient ascent algorithm to design hybrid precoders. Simulation results show that

1) the proposed algorithm converges to a near globally optimal solution from arbitrary initial points; 2) the performance of mixed-CSI-based hybrid precoding is very close to that of instantaneous-CSI-based hybrid precoding. 3) the statistical-CSI-based hybrid precoding can achieve higher energy efficiency than the fully-connected hybird precoding.

B. Notations

The following notations are adopted throughout the paper: Boldface lowercase letters, boldface uppercase letters, and calligraphic letters are used to denote vectors, matrices and sets, respectively. The real and complex number fields are denoted by $\mathbb{R}$ and $\mathbb{C}$, respectively. The superscripts $(\cdot)^{\mathrm{T}}$, $(\cdot)^*$ and $(\cdot)^{\mathrm{H}}$ stand for transpose, conjugate, and conjugate transpose operations, respectively. $\mathrm{tr}(\cdot)$ is the trace of a matrix; $\|\cdot\|$ denotes the Euclidean norm of a vector; $\|\cdot\|_F$ represents the Frobenius norm of a matrix; $E_{\mathbf{x}}(\cdot)$ represents the statistical expectation with respect to $\mathbf{x}$; $\mathbf{X}_{kl}$ represents the (k,l)-th element of $\mathbf{X}$; $\mathbf{I}$ and $\mathbf{0}$ denote an identity matrix and a zero matrix, respectively, with appropriate dimensions; $\circ$ represents the Hadamard matrix product; $\mathcal{I}(\cdot)$ represents the mutual information; $\Re$ and $\Im$ are the real and imaginary parts of a complex value; $\log(\cdot)$ is used for the base two logarithm.

II. SYSTEM AND CHANNEL MODELS

In this section, we present system and channel models for mmWave MIMO systems.

A. System Model

Consider a point-to-point mmWave MIMO system, where a transmitter with N_{t} antennas sends N_{s} data streams to a receiver with N_{r} antennas. The number of RF chains at the transmitter is N_{rf}, which satisfies $N_{\mathrm{s}} \leq N_{\mathrm{rf}} \leq N_{\mathrm{t}}$. We consider the hybrid precoding scheme, where N_{s} data streams are first precoded using a digital precoder, and then shaped by an analog precoder. The received baseband signal $\mathbf{y} \in \mathbb{C}^{N_{\mathrm{r}} \times 1}$ can be written as

$$\mathbf{y} = \mathbf{HFBx} + \mathbf{n} \tag{1}$$

where $\mathbf{H} \in \mathbb{C}^{N_{\mathrm{r}} \times N_{\mathrm{t}}}$ is the mmWave channel matrix; $\mathbf{F} \in \mathbb{C}^{N_{\mathrm{t}} \times N_{\mathrm{rf}}}$ is the analog precoder; $\mathbf{B} \in \mathbb{C}^{N_{\mathrm{rf}} \times N_{\mathrm{s}}}$ is the digital precoder; $\mathbf{x} \in \mathbb{C}^{N_{\mathrm{s}} \times 1}$ is the input data vector and $\mathbf{n} \in \mathbb{C}^{N_{\mathrm{r}} \times 1}$ is the independent and identically distributed (i.i.d.) complex Gaussian noise with zero-mean and covariance $\sigma^2 \mathbf{I}$. To simplify our system model, we omit the analog and digital combiner, which can be designed similarly as the hybrid precoder.

In this paper, the analog precoder $\mathbf{F}$ is implemented by a dynamic phase shifter subarray, where each RF chain is connected to a dynamic subset of transmit antennas. Let $\mathcal{S}_j$ denote the collection of transmit antennas connected to jth RF chain. We partition N_{t} transmit antennas into N_{rf}

subsets satisfying

$$\mathcal{S} = \left\{ \{\mathcal{S}_j\}_{j=1}^{N_{\rm rf}} \left| \bigcup_{j=1}^{N_{\rm rf}} \mathcal{S}_j = \{1, 2, \dots, N_{\rm t}\}, \right. \right.$$
$$\left. \mathcal{S}_j \cap \mathcal{S}_k = \emptyset, \ \forall j \neq k \right\}. \quad (2)$$

Since each RF chain can be connected to different number of antennas, the cardinalities of $\{\mathcal{S}_j\}_{j=1}^{N_{\rm rf}}$ are different. In addition, if the ith transmit antenna is connected to the jth RF chain, i.e.,$i \in \mathcal{S}_j$, the (i, j)th entry of $\mathbf{F}$ has unit modulus, otherwise it is zero. Therefore, the constraints on $\mathbf{F}$ can be expressed by

$$|\mathbf{F}_{ij}| = \mathbf{1}_{\mathcal{S}_j}(i), \quad \forall (i, j) \quad (3)$$

where $\mathbf{1}_{\mathcal{S}_j}(i)$ is the indicator function:

$$\mathbf{1}_{\mathcal{S}_j}(i) = \begin{cases} 1 & \text{if } i \in \mathcal{S}_j \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

The transmitted signal is restricted by a total power constraint P:

$$E_{\mathbf{x}} \|\mathbf{FBx}\|^2 = \text{tr}\left(\mathbf{B}^H \mathbf{F}^H \mathbf{FB}\right) \leq P. \quad (5)$$

To decouple $\mathbf{F}$ and $\mathbf{B}$ in coupled power constraint (5), we consider the following change of variables:

$$\bar{\mathbf{F}} = \mathbf{F}(\mathbf{F}^H \mathbf{F})^{-\frac{1}{2}} \quad (6)$$
$$\bar{\mathbf{B}} = (\mathbf{F}^H \mathbf{F})^{\frac{1}{2}} \mathbf{B}. \quad (7)$$

Then the power constraint in (5) becomes

$$\mathcal{B} = \left\{ \bar{\mathbf{B}} \left| \text{tr}\left(\bar{\mathbf{B}}^H \bar{\mathbf{B}}\right) \leq P \right. \right\} \quad (8)$$

and the constraints on $\bar{\mathbf{F}}$ can be expressed by

$$\mathcal{F} = \left\{ \bar{\mathbf{F}} \left| |\bar{\mathbf{F}}_{ij}| = |\mathcal{S}_j|^{-\frac{1}{2}} \mathbf{1}_{\mathcal{S}_j}(i), \ \forall (i, j) \right. \right\}. \quad (9)$$

Furthermore, by plugging $\bar{\mathbf{F}}$ and $\bar{\mathbf{B}}$ into the system model in (1), we have

$$\mathbf{y} = \mathbf{H}\bar{\mathbf{F}}\bar{\mathbf{B}}\mathbf{x} + \mathbf{n}. \quad (10)$$

Combining (8) and (10), we observe that $\mathbf{H}\bar{\mathbf{F}}$ and $\bar{\mathbf{B}}$ can be regarded as the effective channel and precoder for typical MIMO Gaussian channels, respectively. Since there exists a one-to-one mapping between $(\mathbf{F}, \mathbf{B})$ and $(\bar{\mathbf{F}}, \bar{\mathbf{B}})$, we will focus on designing the effective analog and digital precoders $(\bar{\mathbf{F}}, \bar{\mathbf{B}})$ throughout the rest of this paper.

B. Channel Model

The mmWave MIMO channel is characterized by a standard multi-path model [29, ch. 7.3.2]:

$$\mathbf{H} = \sqrt{\frac{N_{\rm r} N_{\rm t}}{L}} \sum_{\ell=1}^{L} \gamma_\ell \mathbf{a}(\theta_{\rm r,\ell}) \mathbf{a}(\theta_{\rm t,\ell})^H \quad (11)$$

where L denotes the number of physical propagation paths between the transmitter and the receiver; γ_ℓ represents the complex gain of the ℓth propagation path; We assume that

γ_ℓ are i.i.d. complex Gaussian distributed with zero-mean and unit-variance; $\mathbf{a}(\theta_{\rm r,\ell})$ and $\mathbf{a}(\theta_{\rm t,\ell})$ represent the receive and transmit array steering vectors, with $\theta_{\rm r,\ell}$ and $\theta_{\rm t,\ell}$ being the angles of arrival (AOA) and the angles of departure (AOD), respectively. In this paper, the transmitter and receiver adopt uniform linear arrays, whose array steering vector $\mathbf{a}(\theta)$ is given by

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{N}} \left[1, e^{-j\frac{2\pi}{\lambda}d\sin\theta}, \dots, e^{-j\frac{2\pi}{\lambda}d(N-1)\sin\theta} \right]^T \quad (12)$$

where N is the number of antenna element, λ is the wavelength of the carrier frequency and $d = \frac{1}{2}\lambda$ is the antenna spacing.

The channel model in (11) can be rewritten more compactly as

$$\mathbf{H} = \sqrt{\frac{N_{\rm r} N_{\rm t}}{L}} \mathbf{A}_{\rm r} \mathbf{\Gamma} \mathbf{A}_{\rm t}^H \quad (13)$$

where $\mathbf{\Gamma} = \text{diag}(\gamma_1, \dots, \gamma_L)$; $\mathbf{A}_{\rm r} \in \mathbb{C}^{N_{\rm r} \times L}$ and $\mathbf{A}_{\rm t} \in \mathbb{C}^{N_{\rm t} \times L}$ are stacked array steering vectors of AOA and AOD respectively, given by

$$\mathbf{A}_{\rm r} = \left[\mathbf{a}(\theta_{\rm r,1}), \dots, \mathbf{a}(\theta_{\rm r,L}) \right] \quad (14)$$
$$\mathbf{A}_{\rm t} = \left[\mathbf{a}(\theta_{\rm t,1}), \dots, \mathbf{a}(\theta_{\rm t,L}) \right]. \quad (15)$$

This work assumes that the small scale fading $\mathbf{\Gamma}$ varies rapidly while the variation of angle information $\mathbf{A}_{\rm r}$ and $\mathbf{A}_{\rm t}$ is slow [30]. Since the angle information changes slowly, we further assume that the transmitter can obtain statistical CSI through feedback, i.e., the transmitter knows $\mathbf{A}_{\rm r}$ and $\mathbf{A}_{\rm t}$.

III. PROBLEM FORMULATION

For mmWave MIMO systems, it may not be practical to obtain the instantaneous CSI by conventional channel estimation techniques because 1) the channel matrix measured in the baseband depends on the choice of analog precoder; 2) the training blocks may be prohibitively long due to the large bandwidth and low signal-to-noise ratio (SNR). To mitigate this difficulty, we propose new formulations in which analog and/or digital precoders are designed under statistical CSI.

A. Statistical-CSI-Based Formulation

We assume that the transmitter has the knowledge of statistical CSI, including $\mathbf{A}_{\rm r}$, $\mathbf{A}_{\rm t}$ and the distribution of $\mathbf{\Gamma}$. Then we design the analog and digital precoder to maximize the average mutual information. Suppose each entry of the input data vector $\mathbf{x}$ is uniformly distributed from a given constellation set with cardinality M. The average mutual information between $\mathbf{x}$ and $\mathbf{y}$ is given by

$$E_{\mathbf{H}} \mathcal{I}(\mathbf{x}; \mathbf{y} | \mathbf{H}) \quad (16)$$

where $\mathcal{I}(\mathbf{x}; \mathbf{y} | \mathbf{H})$ is the instantaneous mutual information between $\mathbf{x}$ and $\mathbf{y}$ [20]

$$\mathcal{I}(\mathbf{x}; \mathbf{y} | \mathbf{H}) = \log K - \frac{1}{K} \sum_{m=1}^{K} E_{\mathbf{n}} \left\{ \log \sum_{k=1}^{K} \exp(-d_{mk}) \right\}. \quad (17)$$

Here $K = M^{N_s}$ is a constant number, and $d_{mk} = \sigma^{-2}\left(\|\mathbf{H}\mathbf{F}\bar{\mathbf{B}}(\mathbf{x}_m - \mathbf{x}_k) + \mathbf{n}\|^2 - \|\mathbf{n}\|^2\right)$, with $\mathbf{x}_m$ and $\mathbf{x}_k$ being two possible data vectors taken from $\mathbf{x}$. The average mutual information maximization problem can then be formulated as

$$\underset{\{\mathcal{S}_j\}\in\mathcal{S}}{\text{maximize}}\ R(\{\mathcal{S}_j\}) \tag{18}$$

where $R(\{\mathcal{S}_j\})$ is the maximum average mutual information with given partition of subsets, i.e.,

$$R(\{\mathcal{S}_j\}) = \underset{\bar{\mathbf{F}}\in\mathcal{F},\bar{\mathbf{B}}\in\mathcal{B}}{\text{maximize}}\ E_{\mathbf{H}}\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H}). \tag{19}$$

Problem (18) is a combinatorial optimization problem for which finding the optimal solution requires an exhaustive search over all nonempty $\{\mathcal{S}_j\}_{j=1}^{N_{\text{rf}}}$ in $\mathcal{S}$. The total number of combinations is known as Stirling number of the second kind [31] and is given by

$$|\mathcal{S}| = \frac{1}{N_{\text{rf}}!}\sum_{k=0}^{N_{\text{rf}}}(-1)^{N_{\text{rf}}-k}\binom{N_{\text{t}}}{k}k^{N_{\text{t}}}. \tag{20}$$

Then we can rewrite problem (18) as

$$\underset{\ell\in\{1,\ldots,|\mathcal{S}|\}}{\text{maximize}}\ R(\{\mathcal{S}_{j,\ell}\}) \tag{21}$$

where $\{\mathcal{S}_{j,\ell}\}$ represents the ℓth given partition of subsets belonging to $\mathcal{S}$.

Although (21) provides a theoretically possible way for solving problem (18), its computational complexity is prohibitive even for a small number of transmit antennas and RF chains. For example, when $N_{\text{t}} = 16$ and $N_{\text{rf}} = 4$, $|\mathcal{S}|$ is equal to 1.718×10^8, which implies that we need to solve problem (19) over ten million times to obtain the optimal analog and digital precoder.

We propose a new formulation to reduce the computational complexity of problem (18). Recall that $\{\mathcal{S}_j\}_{j=1}^{N_{\text{rf}}}$ represent positions of nonzero entries in $\bar{\mathbf{F}}$, and the role of $\bar{\mathbf{F}}$ is to reshape the effective channel matrix $\mathbf{H}\bar{\mathbf{F}}$. Therefore, we design $\{\mathcal{S}_j\}_{j=1}^{N_{\text{rf}}}$ and the corresponding $\bar{\mathbf{F}}$ such that the average effective channel gain $E_{\mathbf{H}}\|\mathbf{H}\bar{\mathbf{F}}\|_F^2$ is maximized. The dynamic subarray design problem can then be formulated as

$$\underset{\bar{\mathbf{F}}\in\mathcal{F},\{\mathcal{S}_j\}\in\mathcal{S}}{\text{maximize}}\ E_{\mathbf{H}}\text{tr}\left(\bar{\mathbf{F}}^H\mathbf{H}^H\mathbf{H}\bar{\mathbf{F}}\right). \tag{22}$$

We solve problem (22) to obtain its optimal solutions, denoted by $\bar{\mathbf{F}}_{\text{init}}^\star$ and $\{\mathcal{S}_j^\star\}_{j=1}^{N_{\text{rf}}}$. Then we solve problem (19) with given $\{\mathcal{S}_j^\star\}_{j=1}^{N_{\text{rf}}}$ to obtain the optimally effective analog and digital precoders $(\bar{\mathbf{F}}^\star, \bar{\mathbf{B}}^\star)$. Note that since $\bar{\mathbf{F}}_{\text{init}}^\star$ is not obtained by maximizing the average mutual information, we do not use it directly as the optimally effective analog precoder. However, the solution $\bar{\mathbf{F}}_{\text{init}}^\star$ serves as a good initial point for solving problem (19). Therefore, we first design a low complexity algorithm to solve problem (22), and then design an effective algorithm to solve the hybrid precoding problem (19) with given $\{\mathcal{S}_j^\star\}_{j=1}^{N_{\text{rf}}}$.

B. Mixed-CSI-Based Formulation

The basic idea of mixed CSI based formulation is to design the analog precoder based on statistical CSI, and then estimate the reduced-dimensional effective channel matrix $\mathbf{H}\bar{\mathbf{F}}^\star$, where $\bar{\mathbf{F}}^\star = \mathbf{F}(\mathbf{F}^H\mathbf{F})^{-\frac{1}{2}}$ is the optimally effective analog precoder based on statistical CSI. After that, the transmitter utilizes the *instantaneous* effective channel matrix $\mathbf{H}\bar{\mathbf{F}}^\star$ to design effective digital precoder $\bar{\mathbf{B}}$, and this is a typical MIMO precoding problem. In this case, the burden of channel estimation is greatly reduced because the dimension of $\mathbf{H}\bar{\mathbf{F}}^\star \in \mathbb{C}^{N_{\text{r}}\times N_{\text{rf}}}$ is much smaller than that of $\mathbf{H} \in \mathbb{C}^{N_{\text{r}}\times N_{\text{t}}}$.

Given the instantaneous effective channel matrix $\mathbf{H}\bar{\mathbf{F}}$, the digital precoding problem can be expressed by

$$C(\mathbf{H}\bar{\mathbf{F}}) = \underset{\bar{\mathbf{B}}\in\mathcal{B}}{\text{maximize}}\ \mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H}) \tag{23}$$

where $C(\mathbf{H}\bar{\mathbf{F}})$ is the maximum mutual information under the given effective channel matrix $\mathbf{H}\bar{\mathbf{F}}$. Then the mixed-CSI-based hybrid precoding problem can be formulated as

$$\underset{\bar{\mathbf{F}}\in\mathcal{F},\{\mathcal{S}_j\}\in\mathcal{S}}{\text{maximize}}\ E_{\mathbf{H}}C(\mathbf{H}\bar{\mathbf{F}}). \tag{24}$$

Problem (24) is intractable because it is prohibitive to compute the objective function $E_{\mathbf{H}}C(\mathbf{H}\bar{\mathbf{F}})$. In order to estimate $E_{\mathbf{H}}C(\mathbf{H}\bar{\mathbf{F}})$ at a given point $\bar{\mathbf{F}}$, we need to solve the nonconvex problem (23) thousands of times for randomly generated channel matrix $\mathbf{H}$. To mitigate this difficulty, we replace $E_{\mathbf{H}}C(\mathbf{H}\bar{\mathbf{F}})$ by a computationally efficient bound. Invoke Jensen's inequality, $E_{\mathbf{H}}C(\mathbf{H}\bar{\mathbf{F}})$ can be lower bounded by

$$E_{\mathbf{H}}C(\mathbf{H}\bar{\mathbf{F}}) \geq \underset{\bar{\mathbf{B}}\in\mathcal{B}}{\text{maximize}}\ E_{\mathbf{H}}\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H}). \tag{25}$$

Replacing $E_{\mathbf{H}}C(\mathbf{H}\bar{\mathbf{F}})$ by its lower bound, problem (24) is approximated as

$$\underset{\bar{\mathbf{F}}\in\mathcal{F},\{\mathcal{S}_j\}\in\mathcal{S},\bar{\mathbf{B}}\in\mathcal{B}}{\text{maximize}}\ E_{\mathbf{H}}\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H}) \tag{26}$$

which is exactly the same as problem (18). Then we can use the same procedure to solve this problem, i.e., we first solve problem (22) to obtain $\{\mathcal{S}_j^\star\}_{j=1}^{N_{\text{rf}}}$, and then solve problem (19) with given $\{\mathcal{S}_j^\star\}_{j=1}^{N_{\text{rf}}}$ to obtain the optimally effective analog precoder. Note that although the statistical-CSI-based formulation and the mixed-CSI-based formulation solve the same optimization problem, there is an important difference between them. The optimization variable $\bar{\mathbf{B}}$ in the mixed-CSI-based formulation is just an auxiliary variable made for analog precoder design. After obtaining the optimally effective analog precoder, the real digital precoder should be obtained by solving problem (23).

IV. DYNAMIC SUBARRAY DESIGN

In this section, we propose a low complexity algorithm to solve problem (22). Note that the objective function in problem (22) can be rewritten as

$$E_{\mathbf{H}}\text{tr}\left(\bar{\mathbf{F}}^H\mathbf{H}^H\mathbf{H}\bar{\mathbf{F}}\right) = \frac{N_{\text{r}}N_{\text{t}}}{L}E_{\boldsymbol{\Gamma}}\text{tr}\left(\bar{\mathbf{F}}^H\mathbf{A}_{\text{t}}\boldsymbol{\Gamma}^H\mathbf{A}_{\text{r}}^H\mathbf{A}_{\text{r}}\boldsymbol{\Gamma}\mathbf{A}_{\text{t}}^H\bar{\mathbf{F}}\right)$$
$$= \frac{N_{\text{r}}N_{\text{t}}}{L}\text{tr}\left(\bar{\mathbf{F}}^H\mathbf{A}_{\text{t}}\mathbf{A}_{\text{t}}^H\bar{\mathbf{F}}\right) \tag{27}$$

where the second equality in equation (27) holds because $E_{\boldsymbol{\Gamma}}(\boldsymbol{\Gamma}^H \mathbf{A}_r^H \mathbf{A}_r \boldsymbol{\Gamma}) = \mathbf{I}$. Plugging $\bar{\mathbf{F}} = \mathbf{F}(\mathbf{F}^H \mathbf{F})^{-\frac{1}{2}}$ into equation (28), we obtain the following problem

$$\underset{\mathbf{F}, \{\mathcal{S}_j\}}{\text{maximize}} \ \text{tr}\left[(\mathbf{F}^H \mathbf{F})^{-\frac{1}{2}} \mathbf{F}^H \mathbf{A}_t \mathbf{A}_t^H \mathbf{F}(\mathbf{F}^H \mathbf{F})^{-\frac{1}{2}}\right]$$
$$\text{subject to } |\mathbf{F}_{ij}| = \mathbf{1}_{\mathcal{S}_j}(i), \quad \forall(i,j)$$
$$\{\mathcal{S}_j\} \in \mathcal{S}. \tag{28}$$

It is difficult to solve problem (28) directly because the feasible set of problem (28) is characterized by $\mathbf{F}$ and $\{\mathcal{S}_j\}_{j=1}^{N_{\text{rf}}}$. To address this issue, the following proposition rewrites the feasible set as explicit constraints of $\mathbf{F}$.

Proposition 1: The feasible set of problem (28) can be expressed by

$$|\mathbf{F}_{ij}| \in \{0, 1\}, \quad \forall(i,j)$$
$$\|\mathbf{F}_{i\bullet}\|_0 = 1, \quad \forall i \tag{29}$$

where $\mathbf{F}_{i\bullet}$ denotes the ith row of $\mathbf{F}$, and $\|\cdot\|_0$ represents the total number of nonzero elements in a vector.

Proof: See Appendix. ■

According to Proposition 1, we rewrite problem (28) as

$$\underset{\mathbf{F}}{\text{maximize}} \ \text{tr}\left[(\mathbf{F}^H \mathbf{F})^{-\frac{1}{2}} \mathbf{F}^H \mathbf{A}_t \mathbf{A}_t^H \mathbf{F}(\mathbf{F}^H \mathbf{F})^{-\frac{1}{2}}\right]$$
$$\text{subject to } |\mathbf{F}_{ij}| \in \{0, 1\}, \quad \forall(i,j)$$
$$\|\mathbf{F}_{i\bullet}\|_0 = 1, \quad \forall i. \tag{30}$$

Problem (30) is still intractable due to nonconvex discrete constraints $|\mathbf{F}_{ij}| \in \{0, 1\}$ and $\|\mathbf{F}_{i\bullet}\|_0 = 1$. Therefore, we first drop the constraints and consider the unconstrained problem

$$\underset{\mathbf{F}}{\text{maximize}} \ \text{tr}\left[(\mathbf{F}^H \mathbf{F})^{-\frac{1}{2}} \mathbf{F}^H \mathbf{A}_t \mathbf{A}_t^H \mathbf{F}(\mathbf{F}^H \mathbf{F})^{-\frac{1}{2}}\right]. \tag{31}$$

Problem (31) is a generalized eigenvalue problem, and its optimal solution is given by [15]

$$\mathbf{F} = \mathbf{U_A R} \tag{32}$$

where $\mathbf{U_A} \in \mathbb{C}^{N_t \times N_{\text{rf}}}$ is the left singular vectors of $\mathbf{A}_t$ corresponding to the largest N_{rf} singular values, and $\mathbf{R} \in \mathbb{C}^{N_{\text{rf}} \times N_{\text{rf}}}$ is an arbitrary unitary matrix. Note that when $L < N_{\text{rf}}$, the remaining $N_{\text{rf}} - L$ left singular vectors in $\mathbf{U_A}$ can be chosen arbitrarily as long as $\mathbf{U_A}$ satisfies $\mathbf{U_A}^H \mathbf{U_A} = \mathbf{I}$.

In general, if there exists a unitary matrix $\mathbf{R}$ such that the unconstrained optimal solution $\mathbf{U_A R}$ satisfies (29), then $\mathbf{U_A R}$ is the globally optimal solution of problem (30). However, such $\mathbf{R}$ may not exist and thus we use $\mathbf{U_A R}$ to find a nearby feasible solution. Specifically, consider the following optimization problem

$$\underset{\mathbf{F}, \mathbf{R} \in \mathcal{U}}{\text{minimize}} \ \|\mathbf{F} - \mathbf{U_A R}\|_F^2$$
$$\text{subject to } |\mathbf{F}_{ij}| \in \{0, 1\}, \quad \forall(i,j)$$
$$\|\mathbf{F}_{i\bullet}\|_0 = 1, \quad \forall i \tag{33}$$

where $\mathcal{U}$ denotes the set of unitary matrices. Since the optimization variables $\mathbf{F}$ and $\mathbf{R}$ are separate, we adopt the alternating minimization approach to solve problem (33).

Given $\mathbf{R}$, the optimal $\mathbf{F}$ of problem (33) has a simple closed form solution. Let $j^\star(i) = \text{argmax}_{1 \le j \le N_{\text{rf}}} |[\mathbf{U_A R}]_{ij}|$, then the optimal $\mathbf{F}$ of problem (33) can be expressed by

$$\mathbf{F}_{ij} = \begin{cases} \dfrac{[\mathbf{U_A R}]_{ij}}{|[\mathbf{U_A R}]_{ij}|} & \text{if } j = j^\star(i) \\ 0 & \text{otherwise.} \end{cases} \tag{34}$$

Given $\mathbf{F}$, problem (33) is reduced to an orthogonal procrustes problem

$$\underset{\mathbf{R} \in \mathcal{U}}{\text{minimize}} \ \|\mathbf{F} - \mathbf{U_A R}\|_F^2. \tag{35}$$

Let the singular value decomposition of $\mathbf{Z} = \mathbf{F}^H \mathbf{U_A}$ be

$$\mathbf{Z} = \mathbf{F}^H \mathbf{U_A} = \mathbf{U_Z \Sigma_Z V_Z}^H \tag{36}$$

where $\mathbf{U_Z}$ is a unitary matrix with left singular vectors, $\mathbf{\Sigma_Z}$ is a diagonal matrix with singular values arranged in decreasing order, and $\mathbf{V_Z}$ is another unitary matrix with right singular vectors. Then the optimal solution of problem (35) is given by [32]

$$\mathbf{R} = \mathbf{V_Z U_Z}^H. \tag{37}$$

Combining (34) and (37), we propose a simple alternating minimization algorithm to solve problem (33) and obtain the corresponding near optimal partition of subsets $\{\mathcal{S}_j\}_{j=1}^{N_{\text{rf}}}$. The details of this algorithm is summarized in Algorithm 1.

Algorithm 1 Dynamic Subarray Design

---

1. Given the stacked array steering vectors of AOD $\mathbf{A}_t$. Compute $\mathbf{A}_t$'s left singular vectors $\mathbf{U_A}$ and generate an arbitrary initial unitary matrix $\mathbf{R}$.
2. **While** the stopping criterion is not satisfied
   - Given $\mathbf{R}$, solve problem (33) to obtain the optimal $\mathbf{F}$ in (34).
   - Given $\mathbf{F}$, solve problem (33) to obtain the optimal $\mathbf{R}$ in (37).
3. Return $\bar{\mathbf{F}}_{\text{init}}^\star = \mathbf{F}(\mathbf{F}^H \mathbf{F})^{-\frac{1}{2}}$ and the corresponding $\{\mathcal{S}_j^\star\}$.

We conclude this section with several remarks on Algorithm 1:

- The convergence of Algorithm 1 is guaranteed because the objective function $\|\mathbf{F} - \mathbf{U_A R}\|_F^2$ is bounded, and it is decreasing in each iteration.
- Since problem (33) is a nonconvex problem, the solution obtained by Algorithm 1 depends on the initial unitary matrix $\mathbf{R}$. Therefore, we can run Algorithm 1 several times with different initial $\mathbf{R}$, and then choose the solution corresponding to the largest $\|\mathbf{A}_t^H \bar{\mathbf{F}}_{\text{init}}^\star\|_F^2$.
- When $\bar{\mathbf{F}}_{\text{init}}^\star = \mathbf{F}(\mathbf{F}^H \mathbf{F})^{-\frac{1}{2}}$ is determined, the corresponding $\{\mathcal{S}_j^\star\}$ is given by

$$\mathcal{S}_j^\star = \left\{ i \ \middle| \ |[\bar{\mathbf{F}}_{\text{init}}^\star]_{ij}| \ne 0 \right\}, \quad j = 1, \dots, N_{\text{rf}}.$$

V. HYBRID PRECODING WITH FINITE-ALPHABET INPUTS

In this section, we first derive the lower bound for the average mutual information $E_{\mathbf{H}} \mathcal{I}(\mathbf{x}; \mathbf{y}|\mathbf{H})$, and then propose an effective algorithm to design analog and digital precoders.

A. Lower Bound for Average Mutual Information

It is difficult to compute and optimize the average constellation-constrained mutual information directly because both $E_{\mathbf{H}}\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H})$ and its gradient have no closed form expressions. To estimate $E_{\mathbf{H}}\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H})$ as well as its gradient, we need to use Monte Carlo method and/or numerical integral, whose computational complexity are prohibitively high.

This difficulty can be partially mitigated by the following proposition, which provides the lower bound of $E_{\mathbf{H}}\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H})$ in closed form.

Proposition 2: The average constellation-constrained mutual information of mmWave MIMO channels can be lower bounded by

$$L(\bar{\mathbf{F}},\bar{\mathbf{B}}) = \log K - N_r\left(\frac{1}{\ln 2}-1\right)$$
$$-\frac{1}{K}\sum_{m=1}^{K}\log\sum_{k=1}^{K}\det\left[\mathbf{I}+\left(\mathbf{A}_r^H\mathbf{A}_r\right)^T\circ\mathbf{W}_{mk}\right]^{-1} \tag{38}$$

where

$$\mathbf{W}_{mk} = \frac{N_r N_t}{2\sigma^2 L}\mathbf{A}_t^H\bar{\mathbf{F}}\bar{\mathbf{B}}(\mathbf{x}_m-\mathbf{x}_k)(\mathbf{x}_m-\mathbf{x}_k)^H\bar{\mathbf{B}}^H\bar{\mathbf{F}}^H\mathbf{A}_t. \tag{39}$$

Proof: See Appendix. ∎

The computational complexity of the lower bound $L(\bar{\mathbf{F}},\bar{\mathbf{B}})$ is still very high because it needs to calculate the determinant K^2 times. For example, when we adopt 16QAM modulation ($M = 16$) and the number of data streams N_s is 4, K^2 is equal to 4.295×10^9. To further reduce the complexity, we notice that the receive steering vectors are asymptotically orthogonal to each other when the number of receive antennas N_r approaches infinity, i.e., $\lim_{N_r\to\infty}\mathbf{A}_r^H\mathbf{A}_r = \mathbf{I}$. Based on this observation, we derive a low complexity approximation of $L(\bar{\mathbf{F}},\bar{\mathbf{B}})$ in the following proposition.

Proposition 3: The lower bound $L(\bar{\mathbf{F}},\bar{\mathbf{B}})$ can be approximated by

$$L_A(\bar{\mathbf{F}},\bar{\mathbf{B}}) = \log K - N_r\left(\frac{1}{\ln 2}-1\right)$$
$$-\frac{1}{K}\sum_{m=1}^{K}\log\sum_{k=1}^{K}\prod_{\ell=1}^{L}\left(1+\frac{N_r N_t}{2\sigma^2 L}|\beta_{mk\ell}|^2\right)^{-1} \tag{40}$$

where $\beta_{mk\ell} = \mathbf{a}(\theta_{t,\ell})^H\bar{\mathbf{F}}\bar{\mathbf{B}}(\mathbf{x}_m-\mathbf{x}_k)$, with $\mathbf{a}(\theta_{t,\ell})$ being the ℓth column of $\mathbf{A}_t$. In addition, the limit of $L_A(\bar{\mathbf{F}},\bar{\mathbf{B}})$ is $L(\bar{\mathbf{F}},\bar{\mathbf{B}})$ as N_r approaches infinity.

Proof: See Appendix. ∎

The accuracy and computational complexity of the lower bound and its approximation will be shown in Fig. 1 and Table 1 in the simulation result section.

B. Hybrid Precoding Design

In this section, we solve the hybrid precoding problem (19) with given $\{\mathcal{S}_j^\star\}_{j=1}^{N_{rf}}$ obtained by Algorithm 1. First, by replacing the average mutual information $E_{\mathbf{H}}\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H})$

with the approximated lower bound $L_A(\bar{\mathbf{F}},\bar{\mathbf{B}})$, problem (19) can be approximated as

$$\underset{\bar{\mathbf{F}},\bar{\mathbf{B}}}{\text{maximize}}\ L_A(\bar{\mathbf{F}},\bar{\mathbf{B}})$$
$$\text{subject to}\ |\bar{\mathbf{F}}_{ij}| = |\mathcal{S}_j^\star|^{-\frac{1}{2}}\mathbf{1}_{\mathcal{S}_j}(i),\quad\forall(i,j)$$
$$\text{tr}\left(\bar{\mathbf{B}}^H\bar{\mathbf{B}}\right)\leq P. \tag{41}$$

Note that the constraint $|\bar{\mathbf{F}}_{ij}| = |\mathcal{S}_j^\star|^{-\frac{1}{2}}\mathbf{1}_{\mathcal{S}_j}(i)$ implies that only the phase of nonzero $|\bar{\mathbf{F}}_{ij}|$ can be changed. Therefore, instead of using $\bar{\mathbf{F}}$ as the optimization variable, it is more convenient to optimize the phase of nonzero entries in $\bar{\mathbf{F}}$. Define the phase matrix $\mathbf{\Phi}$ as

$$\mathbf{\Phi}_{ij} = \angle\bar{\mathbf{F}}_{ij}\mathbf{1}_{\mathcal{S}_j^\star}(i),\quad\forall(i,j) \tag{42}$$

where $\angle\bar{\mathbf{F}}_{ij}$ represents the phase of $\bar{\mathbf{F}}_{ij}$. Then $\bar{\mathbf{F}}$ can be expressed as

$$\bar{\mathbf{F}}_{ij} = |\mathcal{S}_j^\star|^{-\frac{1}{2}}\exp(\jmath\mathbf{\Phi}_{ij})\mathbf{1}_{\mathcal{S}_j^\star}(i),\quad\forall(i,j). \tag{43}$$

Using $\mathbf{\Phi}$ as the optimization variable and defining a new function $R(\mathbf{\Phi},\bar{\mathbf{B}}) \triangleq L_A(\bar{\mathbf{F}}(\mathbf{\Phi}),\bar{\mathbf{B}})$, problem (41) can be rewritten as

$$\underset{\mathbf{\Phi},\bar{\mathbf{B}}}{\text{maximize}}\ R(\mathbf{\Phi},\bar{\mathbf{B}})$$
$$\text{subject to}\ \text{tr}\left(\bar{\mathbf{B}}^H\bar{\mathbf{B}}\right) = P. \tag{44}$$

Here we express the power constraint as $\text{tr}\left(\bar{\mathbf{B}}^H\bar{\mathbf{B}}\right) = P$ because $R(\mathbf{\Phi},\bar{\mathbf{B}})$ is monotonically increasing with respect to $\|\bar{\mathbf{B}}\|_F^2$. Then we provide the gradient of $R(\mathbf{\Phi},\bar{\mathbf{B}})$ in the following proposition, which forms the foundation for solving problem (44).

Proposition 4: The gradient of $R(\mathbf{\Phi},\bar{\mathbf{B}})$ with respect to $\bar{\mathbf{B}}$ and $\mathbf{\Phi}$ are given by

$$\nabla_{\bar{\mathbf{B}}}R(\mathbf{\Phi},\bar{\mathbf{B}}) = \sum_{\ell=1}^{L}\bar{\mathbf{F}}^H\mathbf{a}(\theta_{t,\ell})\mathbf{a}(\theta_{t,\ell})^H\bar{\mathbf{F}}\bar{\mathbf{B}}\mathbf{E}_\ell$$

$$\nabla_{\mathbf{\Phi}}R(\mathbf{\Phi},\bar{\mathbf{B}}) = 2\sum_{\ell=1}^{L}\Im\left[\bar{\mathbf{F}}^H\mathbf{a}(\theta_{t,\ell})\mathbf{a}(\theta_{t,\ell})^H\bar{\mathbf{F}}\bar{\mathbf{B}}\mathbf{E}_\ell\bar{\mathbf{B}}^H\circ\bar{\mathbf{F}}^*\right] \tag{45}$$

where

$$\mathbf{E}_\ell = \frac{1}{\ln(2)\cdot K}\sum_{m,k}\zeta_{mk\ell}(\mathbf{x}_m-\mathbf{x}_k)(\mathbf{x}_m-\mathbf{x}_k)^H \tag{46}$$

with

$$\zeta_{mk\ell} = \left(\frac{2\sigma^2 L}{N_r N_t}+|\beta_{mk\ell}|^2\right)^{-1}\cdot\prod_{\ell=1}^{L}\left(1+\frac{N_r N_t}{2\sigma^2 L}|\beta_{mk\ell}|^2\right)^{-1}$$
$$\cdot\left[\sum_{k=1}^{K}\prod_{\ell=1}^{L}\left(1+\frac{N_r N_t}{2\sigma^2 L}|\beta_{mk\ell}|^2\right)^{-1}\right]^{-1}.$$

Proof: See Appendix. ∎

We propose a manifold-based gradient ascent algorithm to optimize $\bar{\mathbf{\Phi}}$ and $\bar{\mathbf{B}}$ simultaneously using the gradient information. At the kth iteration, the algorithm updates the current solution $(\mathbf{\Phi}_k,\bar{\mathbf{B}}_k)$ to $(\mathbf{\Phi}_{k+1},\bar{\mathbf{B}}_{k+1})$ by the following rules

$$\mathbf{\Phi}_{k+1} = \mathbf{\Phi}_k + \rho_k\nabla_{\mathbf{\Phi}}R(\mathbf{\Phi}_k,\bar{\mathbf{B}}_k)$$
$$\bar{\mathbf{B}}_{k+1} = \text{Proj}\left[\bar{\mathbf{B}}_k + \rho_k\text{grad}_{\bar{\mathbf{B}}}R(\mathbf{\Phi}_k,\bar{\mathbf{B}}_k)\right] \tag{47}$$

where $\rho_k > 0$ is the stepsize, $\mathrm{Proj}\left[\bar{\mathbf{B}}_k\right] = P^{\frac{1}{2}}\|\bar{\mathbf{B}}_k\|_F^{-1}\bar{\mathbf{B}}_k$, and $\mathrm{grad}_{\bar{\mathbf{B}}}R(\mathbf{\Phi},\bar{\mathbf{B}})$ is the gradient of $R(\mathbf{\Phi},\bar{\mathbf{B}})$ on the following (sphere) manifold

$$\mathcal{M} = \left\{\bar{\mathbf{B}}\big|\mathrm{tr}\left(\bar{\mathbf{B}}^H\bar{\mathbf{B}}\right) = P\right\}. \tag{48}$$

Based on the definition, $\mathrm{grad}_{\bar{\mathbf{B}}}R(\mathbf{\Phi},\bar{\mathbf{B}})$ can be computed by projecting $\nabla_{\bar{\mathbf{B}}}R(\mathbf{\Phi},\bar{\mathbf{B}})$ onto the tangent space $T_{\bar{\mathbf{B}}}\mathcal{M}$ at $\bar{\mathbf{B}}$, where $T_{\bar{\mathbf{B}}}\mathcal{M}$ is given by

$$T_{\bar{\mathbf{B}}}\mathcal{M} = \left\{\bar{\mathbf{X}}\big|\mathrm{tr}\left(\bar{\mathbf{X}}^H\bar{\mathbf{B}} + \bar{\mathbf{B}}^H\bar{\mathbf{X}}\right) = 0\right\}. \tag{49}$$

Then $\mathrm{grad}_{\bar{\mathbf{B}}}R(\mathbf{\Phi},\bar{\mathbf{B}})$ can be expressed by

$$\mathrm{grad}_{\bar{\mathbf{B}}}R(\mathbf{\Phi},\bar{\mathbf{B}}) = \underset{\bar{\mathbf{X}}\in T_{\bar{\mathbf{B}}}\mathcal{M}}{\mathrm{argmax}} \left\|\bar{\mathbf{X}} - \nabla_{\bar{\mathbf{B}}}R\right\|_F^2. \tag{50}$$

Using the standard Lagrangian multiplier method, the closed form solution of problem (50) is given by

$$\mathrm{grad}_{\bar{\mathbf{B}}}R(\mathbf{\Phi},\bar{\mathbf{B}}) = \nabla_{\bar{\mathbf{B}}}R(\mathbf{\Phi},\bar{\mathbf{B}}) - \frac{\Re\mathrm{tr}\left[(\nabla_{\bar{\mathbf{B}}}R)^H\bar{\mathbf{B}}\right]}{P}\bar{\mathbf{B}}. \tag{51}$$

After obtaining the ascent direction, we need to determine the stepsize ρ_k such that the objective function $R(\mathbf{\Phi},\bar{\mathbf{B}})$ is increasing in each iteration. We propose a modified backtracking line search method, which is usually more efficient than the classic backtracking line search [33]. The main idea is to use ρ_{k-1} as the initial guess of ρ_k, and then either increases or decreases it to find the largest ρ_k such that

$$\begin{aligned} f(\rho_k) &\triangleq R(\mathbf{\Phi}_{k+1},\bar{\mathbf{B}}_{k+1}) - R(\mathbf{\Phi}_k,\bar{\mathbf{B}}_k) \\ &\quad - \rho_k\beta_{\mathrm{ga}}\big(\|\nabla_{\mathbf{\Phi}}R(\mathbf{\Phi}_k,\bar{\mathbf{B}}_k)\|_F^2 \\ &\quad + \|\mathrm{grad}_{\bar{\mathbf{B}}}R(\mathbf{\Phi}_k,\bar{\mathbf{B}}_k)\|_F^2\big) \geq 0 \end{aligned}$$

where $\beta_{\mathrm{ga}} \in [0, 0.5]$ is a constant to control the stepsize. Specifically, the stepsize ρ_k is set as

$$\rho_k = \begin{cases} 2^{K_1-1} \cdot \rho_{k-1} & \text{if } f(\rho_{k-1}) \geq 0 \\ \left(\frac{1}{2}\right)^{K_2} \cdot \rho_{k-1} & \text{if } f(\rho_{k-1}) < 0 \end{cases} \tag{52}$$

where $K_1 \geq 0$ is the smallest integer such that $f(2^{K_1}\rho_{k-1}) < 0$, and $K_2 \geq 0$ is the smallest integer such that $f([\frac{1}{2}]^{K_2}\rho_{k-1}) \geq 0$. The details of our proposed manifold-based gradient ascent algorithm is summarized in Algorithm 2.

VI. SIMULATION RESULTS

We provide several examples in this section to illustrate the relationship and the computational complexity comparison between average mutual information and its lower bound as well as the lower bound approximation. We also show the convergence of the proposed hybrid precoding algorithm and the efficacy of the designed hybrid precoders. For convenience, we rewrite the angles of arrival $\{\theta_{r,\ell}\}_{\ell=1}^L$ as a vector $\boldsymbol{\theta}_r$, whose ℓth element corresponds to $\theta_{r,\ell}$. Similarly, the angles of departure $\{\theta_{t,\ell}\}_{\ell=1}^L$ can be expressed by $\boldsymbol{\theta}_t$. The angles of arrival follow the Laplacian distribution with a fixed or uniformly distributed mean angle $\bar{\theta}_r$, and a constant angular spread (standard deviation) of $\frac{\pi}{18}$. The angles of departure follow the Laplacian distribution with a fixed mean angle $\bar{\theta}_t$, and an angular spread of $\frac{\pi}{18}$.

Algorithm 2 Manifold-Based Gradient Ascent Algorithm

1. Given $\{\mathcal{S}_j^\star\}_{j=1}^{N_{\mathrm{rf}}}$ (obtained by Algorithm 1), $\mathbf{\Phi}_0$ and $\bar{\mathbf{B}}_0$. Set $\rho_0 = 2$, $\beta_{\mathrm{ga}} = 0.4$, and $\epsilon = 10^{-4}$.
2. For $k = 0, 1, 2, \ldots$ (outer iterations)
   - Compute the gradient of $R(\mathbf{\Phi},\bar{\mathbf{B}})$ with respect to $\mathbf{\Phi}$ and $\bar{\mathbf{B}}$ at $(\mathbf{\Phi}_k,\bar{\mathbf{B}}_k)$ by (45). Then use $\nabla_{\bar{\mathbf{B}}}R(\mathbf{\Phi}_k,\bar{\mathbf{B}}_k)$ to compute the gradient of $R(\mathbf{\Phi},\bar{\mathbf{B}})$ at $\bar{\mathbf{B}}_k$ on the shere manifold by (51).
   - If $\|\nabla_{\mathbf{\Phi}}R(\mathbf{\Phi}_k,\bar{\mathbf{B}}_k)\|_F^2 + \|\mathrm{grad}_{\bar{\mathbf{B}}}R(\mathbf{\Phi}_k,\bar{\mathbf{B}}_k)\|_F^2 < \epsilon$, stop.
   - Utilize the modified backtracking line search to compute the stepsize ρ_k via (52).
   - Update $(\mathbf{\Phi}_k,\bar{\mathbf{B}}_k)$ to $(\mathbf{\Phi}_{k+1},\bar{\mathbf{B}}_{k+1})$ by

$$\mathbf{\Phi}_{k+1} = \mathbf{\Phi}_k + \rho_k\nabla_{\mathbf{\Phi}}R(\mathbf{\Phi}_k,\bar{\mathbf{B}}_k)$$
$$\bar{\mathbf{B}}_{k+1} = \mathrm{Proj}\left[\bar{\mathbf{B}}_k + \rho_k\mathrm{grad}_{\bar{\mathbf{B}}}R(\mathbf{\Phi}_k,\bar{\mathbf{B}}_k)\right].$$

A. Example 1: Average Mutual Information and Lower Bound

This example is utilized to show that 1) the lower bound $L(\bar{\mathbf{F}},\bar{\mathbf{B}})$ plus a constant is a very accurate approximation to the average mutual information; 2) the lower bound approximation $L_A(\bar{\mathbf{F}},\bar{\mathbf{B}})$ plus a constant is also a good approximation to the average mutual information; 3) the computational complexity of $L_A(\bar{\mathbf{F}},\bar{\mathbf{B}})$ is a few orders of magnitudes lower than that of the average mutual information.

We begin with the consideration of limits of the average mutual information. When the noise power σ^2 approaches 0 and $+\infty$, the limits are given by

$$\lim_{\sigma^2 \to 0} E_{\mathbf{H}}\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H}) = \log K \tag{53}$$
$$\lim_{\sigma^2 \to +\infty} E_{\mathbf{H}}\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H}) = s0. \tag{54}$$

At the same time, the limits of $L(\bar{\mathbf{F}},\bar{\mathbf{B}})$ are given by

$$\lim_{\sigma^2 \to 0} L(\bar{\mathbf{F}},\bar{\mathbf{B}}) = \log K - N_r\left(\frac{1}{\ln(2)} - 1\right) \tag{55}$$

$$\lim_{\sigma^2 \to +\infty} L(\bar{\mathbf{F}},\bar{\mathbf{B}}) = -N_r\left(\frac{1}{\ln(2)} - 1\right) \tag{56}$$

which imply that a constant gap $N_r\left(\frac{1}{\ln(2)} - 1\right)$ exists between the average mutual information $E_{\mathbf{H}}\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H})$ and its lower bound $L(\bar{\mathbf{F}},\bar{\mathbf{B}})$ at low and high SNR regimes. Similarly, the same constant gap $N_r\left(\frac{1}{\ln(2)} - 1\right)$ exists between $E_{\mathbf{H}}\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H})$ and $L_A(\bar{\mathbf{F}},\bar{\mathbf{B}})$. Since the optimized hybrid precoders will remain unchanged by adding a constant value to the objective function, we demonstrate that the lower bound $L(\bar{\mathbf{F}},\bar{\mathbf{B}})$ and its approximation $L_A(\bar{\mathbf{F}},\bar{\mathbf{B}})$ plus a constant serve as good approximations to the average mutual information.

We consider a mmWave MIMO system with $N_r = 32$, $N_t = 64$, $N_{\mathrm{rf}} = 4$ and $N_s = 4$. The number of physical propagation paths is set as $L = 6$, and the SNR is defined as $\mathrm{SNR} = \frac{P}{\sigma^2}$. The input signal is drawn from QPSK modulation. The mean angles of $\boldsymbol{\theta}_r$ and $\boldsymbol{\theta}_t$ are set as $\bar{\theta}_r = \bar{\theta}_t = \frac{\pi}{4}$. Then we generate the angles of arrival and departure, whose realizations

TABLE I
RUNNING TIMES (IN SECS.) VERSUS SNR FOR AVERAGE MUTUAL INFORMATION AND ITS APPROXIMATIONS

SNR(dB)	-35	-30	-25	-20	-15	-10	-5
Average mutual information	7.248s	8.969s	8.728s	8.775s	8.857s	8.918s	8.630s
Lower bound w. const. shift	0.230s	0.159s	0.156s	0.197s	0.151s	0.156s	0.211s
Lower bound approx. w. const. shift	0.028s	0.016s	0.014s	0.016s	0.012s	0.012s	0.020s
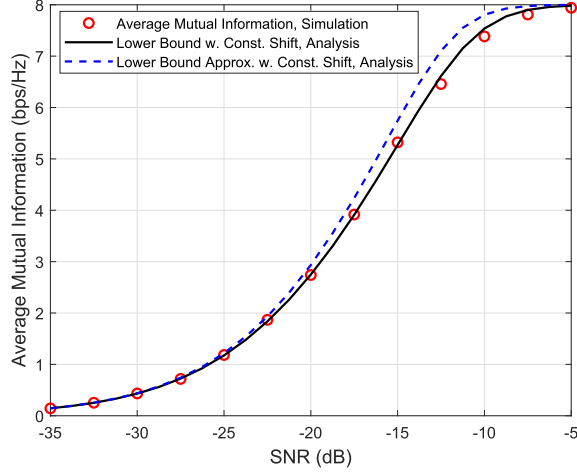
Fig. 1.   Average mutual information with QPSK inputs for mmWave MIMO channels ($N_{\mathrm{r}} = 32$, $N_{\mathrm{t}} = 64$, $N_{\mathrm{rf}} = 4$, $N_{\mathrm{s}} = 4$, $L = 6$).
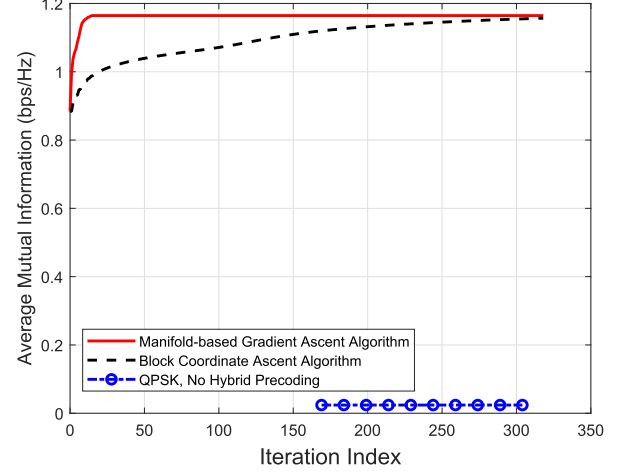

Fig. 2.   Evolution of average mutual information as the hybrid precoders are optimized with the proposed manifold-based gradient ascent and the block coordinate ascent algorithms. The input signal is drawn from QPSK; SNR is -22.5 dB.

are given by

$$\boldsymbol{\theta}_{\mathrm{r}} = [0.6833, 0.5937, 0.5982, 0.5309, 0.7593, 0.7719]^T$$
$$\boldsymbol{\theta}_{\mathrm{t}} = [0.7468, 0.8778, 0.8219, 0.8823, 1.0332, 1.1444]^T. \quad (57)$$

For illustration purpose, the effective analog precoder $\bar{\mathbf{F}}$ is obtained by Algorithm 1, and the effective digital precoder is set as $\bar{\mathbf{B}} = \mathbf{I}$.

The function values and running times for the average mutual information, the lower bound with a constant shift and the lower bound approximation with a constant shift are presented in Fig. 1 and Table I. The simulated curve is obtained by the Monte Carlo method, which computes the average mutual information using 3000 realizations of $\mathbf{H}$ and $\mathbf{n}$. From Fig. 1 and Table I, we have the following remarks:

1) With a constant shift, the lower bound provides a very accurate approximation to the average mutual information in whole SNR regimes.
2) The lower bound approximation plus a constant and the average mutual information match exactly at low and high SNR regimes, and their gap at medium SNR regime is less than 0.5bps/Hz in our case.
3) The lower bound approximation consumes much lower computational time than the average mutual information and its lower bound, thus we design hybrid precoders by maximizing the lower bound approximation.

B. Example 2: Convergence of the Manifold-Based Gradient Ascent Algorithm

In this subsection, we consider a mmWave MIMO system with $N_{\mathrm{r}} = 16$, $N_{\mathrm{t}} = 64$, $N_{\mathrm{rf}} = 4$ and $N_{\mathrm{s}} = 4$. The number of physical propagation paths is set as $L = 6$, and the SNR is given by $\mathrm{SNR} = -22.5$dB. The input signal is drawn

from QPSK modulation. The mean angle of $\boldsymbol{\theta}_{\mathrm{r}}$ is uniformly distributed over $[0, 2\pi]$, i.e., $\bar{\theta}_{\mathrm{r}} \sim \mathrm{unif}(0, 2\pi)$. In contrast, the mean angle of $\boldsymbol{\theta}_{\mathrm{t}}$ is set as $\bar{\theta}_{\mathrm{t}} = \frac{\pi}{3}$. The realizations of $\boldsymbol{\theta}_{\mathrm{r}}$ and $\boldsymbol{\theta}_{\mathrm{t}}$ are given by

$$\boldsymbol{\theta}_{\mathrm{r}} = [4.6448, 4.7492, 4.9337, 4.8962, 5.3448, 4.4681]^T$$
$$\boldsymbol{\theta}_{\mathrm{t}} = [0.8806, 1.4545, 0.8359, 1.1047, 1.2880, 0.8917]^T. \quad (58)$$

The initial point of the effective analog precoder $\bar{\mathbf{F}}_{\mathrm{init}}$ is obtained by Algorithm 1, and the initial point of the effective digital precoder $\bar{\mathbf{B}}_{\mathrm{init}}$ is set as the right singular vectors of $\mathbf{A}_{\mathrm{t}}^H \bar{\mathbf{F}}_{\mathrm{init}}$.

The evolution of the proposed manifold-based gradient ascent algorithm is shown in Fig. 2. For comparison, it also shows the hybrid precoding with block coordinate ascent algorithm, and the average mutual information without hybrid precoding. The block coordinate ascent algorithm solves the hybrid precoding problem (44) by optimizing $\boldsymbol{\Phi}$ and $\bar{\mathbf{B}}$ alternatively with initial point $(\bar{\mathbf{F}}_{\mathrm{init}}, \bar{\mathbf{B}}_{\mathrm{init}})$. The effective analog and digital precoders in no hybrid precoding case are set as

$$\bar{\mathbf{F}} = \sqrt{\frac{N_{\mathrm{rf}}}{N_{\mathrm{t}}}} \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}, \quad \bar{\mathbf{B}} = \mathbf{I}. \quad (59)$$

From Fig. 2, we observe that our proposed manifold-based gradient ascent algorithm converges to 1.165 bps/Hz after 16 iterations while the block coordinate ascent algorithm requires over 320 iterations to approach the same value. Therefore, the proposed manifold-based gradient ascent algorithm is much faster than the block coordinate ascent algorithm. This phenomenon occurs mainly because our proposed algorithm updates $\boldsymbol{\Phi}$ and $\bar{\mathbf{B}}$ simultaneously while the block coordinate
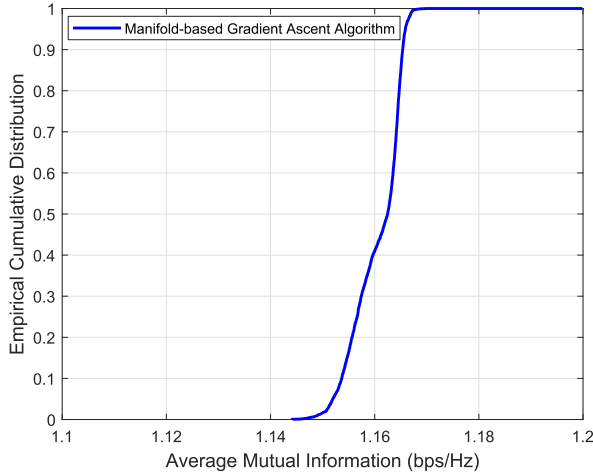
Fig. 3. Empirical cumulative distribution of average mutual information for various initial points. The input signal is drawn from QPSK; SNR is -22.5 dB.

Fig. 4. Average mutual information versus SNR for different scenarios in a mmWave MIMO channel ($N_r = 24$, $N_t = 64$, $N_{rf} = 4$, $N_s = 4$, $L = 8$).

ascent algorithm updates $\mathbf{\Phi}$ and $\bar{\mathbf{B}}$ alternatively. In addition, we also observe that the performance of no hybrid precoding is very poor because we do not exploit any channel state information to design hybrid precoders.

The empirical cumulative distribution of average mutual information for the hybrid precoder from various initial points of the effective analog and digital precoders are further depicted in Fig. 3, which is obtained by generating 3000 random initial points ($\bar{\mathbf{F}}_{init}, \bar{\mathbf{B}}_{init}$). The initial analog precoders $\bar{\mathbf{F}}_{init}$ are obtained by Algorithm 1, whose output depends on the random input matrix $\mathbf{R}$. The initial digital precoders $\bar{\mathbf{B}}_{init}$ are generated with i.i.d. zero-mean unit-variance complex Gaussian entries, and then normalized to satisfy the power constraint. The empirical cumulative distribution curve shows that although the hybrid precoding design with given partition of subsets is a nonconvex problem, our proposed manifold-based gradient ascent algorithm can achieve a near globally optimal solution from arbitrary initial points.

C. Example 3: Performance of Mixed-CSI-Based Hybrid Precoding

In this subsection, we evaluate the performance of mixed-CSI-based hybrid precoding. We consider a mmWave MIMO system with $N_r = 24$, $N_t = 64$, $N_{rf} = 4$ and $N_s = 4$. The number of physical propagation paths is set as $L = 8$. The input signal is drawn from BPSK modulation. The mean angle of $\boldsymbol{\theta}_r$ satisfies $\bar{\theta}_r \sim \text{unif}(0, 2\pi)$, and the mean angle of $\boldsymbol{\theta}_t$ is set as $\bar{\theta}_t = \frac{\pi}{4}$. The realizations of $\boldsymbol{\theta}_r$ and $\boldsymbol{\theta}_t$ are given by

$$\boldsymbol{\theta}_r = [3.921, 3.442, 3.550, 3.449, 3.514, 3.415, 3.314, 3.289]^T$$
$$\boldsymbol{\theta}_t = [0.760, 0.614, 0.674, 0.683, 0.916, 0.749, 0.831, 0.777]^T.$$
$$(60)$$

The mixed-CSI-based hybrid precoding utilizes channel statistics to design the effective analog precoder $\bar{\mathbf{F}}$, and then design the effective digital precoder based on the instantaneous CSI. To evaluate the average mutual information, we generate $N = 1500$ independent samples of the channel matrix

$$\mathbf{H}_i = \sqrt{\frac{N_r N_t}{L}} \mathbf{A}_r \mathbf{\Gamma}_i \mathbf{A}_t^H, \quad i = 1, 2, \ldots, N. \quad (61)$$
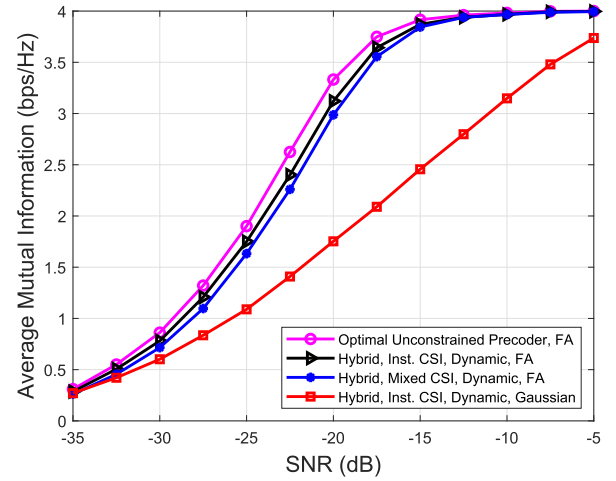
Then we solve the digital precoding problem (23) for each effective channel matrix $\mathbf{H}_i \bar{\mathbf{F}}$. Finally the average mutual information is given by

$$\frac{1}{N} \sum_{i=1}^{N} C(\mathbf{H}_i \bar{\mathbf{F}}) \quad (62)$$

where $C(\mathbf{H}_i \bar{\mathbf{F}})$ is the maximum mutual information for given channel matrix $\mathbf{H}_i \bar{\mathbf{F}}$.

We make comparisons between the mixed-CSI-based hybrid precoding under finite-alphabet (FA) inputs and three interesting scenarios, namely the optimal unconstrained precoder with FA inputs [20], the instantaneous-CSI-based hybrid precoding under FA inputs, and the instantaneous-CSI-based hybrid precoding under Gaussian inputs [15]. All hybrid precoding algorithms are designed for the dynamic subarray structure. The instantaneous-CSI-based hybrid precoding with FA inputs first solve the following dynamic subarray problem

$$\underset{\bar{\mathbf{F}} \in \mathcal{F}, \{\mathcal{S}_j\} \in \mathcal{S}}{\text{maximize}} \ \text{tr}\left(\bar{\mathbf{F}}^H \mathbf{H}^H \mathbf{H} \bar{\mathbf{F}}\right). \quad (63)$$

Note that problem (63) has the same mathematical structure with problem (22), thus we can solve it using Algorithm 1. Then we design analog and digital precoders by maximizing the mutual information with given partition of subsets, i.e., we solve the following optimization problem

$$\underset{\bar{\mathbf{F}} \in \mathcal{F}, \bar{\mathbf{B}} \in \mathcal{B}}{\text{maximize}} \ \mathcal{I}(\mathbf{x}; \mathbf{y} | \mathbf{H}) \quad (64)$$

using the manifold-based gradient ascent algorithm (Algorithm 2).

Fig. 4 demonstrates the average mutual information versus SNR for different scenarios. From Fig. 4, we have the following remarks:

1) The performance of optimal unconstrained precoders is the benchmark for any hybrid precoding schemes, and the proposed hybrid precoding with dynamic subarrays has about 1dB performance loss compared with the optimal unconstrained precoder. Therefore, the hybrid precoding with dynamic subarrays provides a good tradeoff between performance and complexity.

2) The performance gap between our proposed mixed-CSI-based hybrid precoding and the instantaneous-CSI-based hybrid precoding is very small, while the mixed-CSI-based hybrid precoding can greatly reduce the complexity of hybrid precoding design and channel estimation.

3) The mixed-CSI-based hybrid precoding with finite-alphabet inputs can achieve 3.5bps/Hz when $\text{SNR} = -17.5\text{dB}$, while the instantaneous-CSI-based hybrid precoding under Gaussian inputs requires -7.5dB to approach the same value. Therefore, our proposed mixed-CSI-based hybrid precoding has a maximum 10dB gain compared with the instantaneous-CSI-based hybrid precoding under Gaussian inputs. This is mainly because hybrid precoders designed under Gaussian inputs will lead to significant performance loss when applied to systems employing FA.

D. Example 4: Performance of Statistical-CSI-Based Hybrid Precoding

In this subsection, we consider a mmWave MIMO system with $N_r = 4$, $N_t = 64$, $N_{rf} = 4$ and $N_s = 4$. The number of physical propagation paths is set as $L = 5$. The input signal is drawn from QPSK modulation. The mean angle of $\boldsymbol{\theta}_r$ satisfies $\bar{\theta}_r \sim \text{unif}(0, 2\pi)$, and the mean angle of $\boldsymbol{\theta}_t$ is set as $\bar{\theta}_t = \frac{\pi}{4}$. The realizations of $\boldsymbol{\theta}_r$ and $\boldsymbol{\theta}_t$ are given by

$$\boldsymbol{\theta}_r = [0.4186, 0.5499, 0.4839, 0.3135, 0.7505]^T$$
$$\boldsymbol{\theta}_t = [0.9144, 0.7117, 0.7969, 0.8150, 0.6860]^T. \quad (65)$$

We first evaluate the spectral efficiency of the statistical-CSI-based hybrid precoding with dynamic subarrays. We set the fully-connected hybrid precoding under statistical CSI as the benchmark, and then make comparisons between the statistical-CSI-based hybrid precoding with dynamic subarrays and statistical-CSI-based hybrid precoding with fixed subarrays. All hybrid precoding algorithms are designed for FA inputs. The fully-connected hybrid precoder under statistical CSI factorizes the optimal unconstrained precoder into analog and digital precoders [28], and the optimal unconstrained precoder can be obtained by maximizing the lower bound approximation with projected gradient algorithm [33]. The statistical-CSI-based hybrid precoding with fixed subarrays utilizes Algorithm 2 to solve problem (44) with the following given $\{\mathcal{S}_j\}$:

$$\mathcal{S}_j = \big\{(j-1)q + 1, (j-1)q + 2, \ldots, (j-1)q + q\big\}, \quad \forall j \quad (66)$$

where $q = \frac{N_t}{N_{rf}}$. The results in Fig. 5 show that the statistical-CSI-based hybrid precoding with dynamic subarrays has about 1dB performance gain over the statistical-CSI-based hybrid precoding with fixed subarrays in the medium and high SNR regimes.

Then we evaluate the energy efficiency of the statistical-CSI-based hybrid precoding with dynamic subarrays. Based on the energy consumption model in [34], the energy efficiency η is defined as

$$\eta = \frac{E_{\mathbf{H}}\mathcal{I}(\mathbf{x}; \mathbf{y}|\mathbf{H})}{P + N_{rf}P_{rf} + N_{ps}P_{ps}} \quad (67)$$
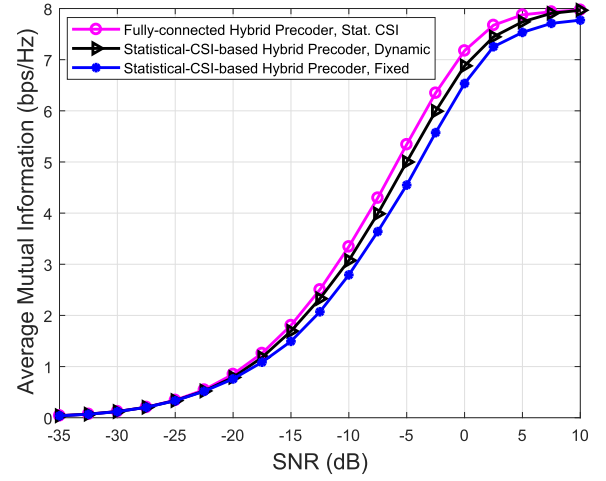

Fig. 5. Average mutual information versus SNR for mmWave MIMO channel with $N_r = 4$, $N_t = 64$, $N_{rf} = 4$, $N_s = 4$, $L = 5$.
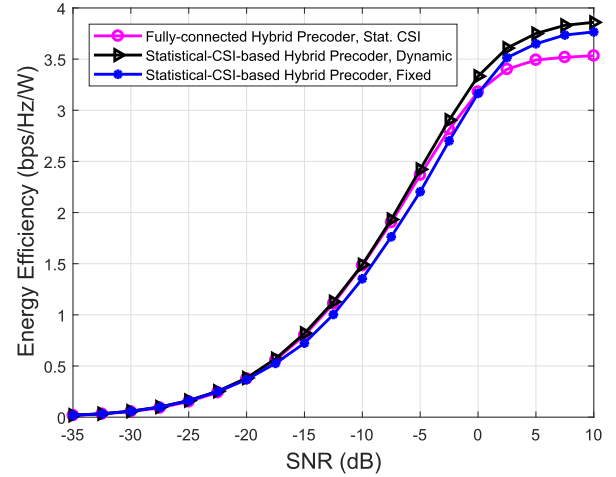

Fig. 6. Energy efficiency versus SNR for mmWave MIMO channel with $N_r = 4$, $N_t = 64$, $N_{rf} = 4$, $N_s = 4$, $L = 5$.

where P is the transmit power, P_{rf} is the power consumed by RF chain, P_{ps} is the power consumed by phase shifter, N_{rf} and N_{ps} are the numbers of required RF chains and phase shifters, respectively. In this paper, we use the practical values $P_{rf} = 250\text{mW}$ [35], $P_{ps} = 1\text{mW}$ [36], and $P = 1\text{W}$ (about 30 dBm) in a small cell transmission scenario [37]. Fig. 6 shows the energy efficiency comparison for the fully-connected hybrid precoding under statistical CSI as well as the statistical-CSI-based hybrid precoding with dynamic and fixed subarrays. We observe that our proposed statistical-CSI-based hybrid precoding with dynamic subarrays outperforms the fully-connected hybrid precoding under statistical CSI in the high SNR regime.

VII. CONCLUSION

In this paper, we have considered the partially-connected hybrid precoding design for millimeter wave (mmWave) multiple-input multiple-output (MIMO) systems with finite-alphabet inputs and dynamic subarrays. The analog and digital precoders are designed using either statistical CSI or

mixed CSI. To simplify the original problem, we have proposed a low complexity algorithm to design the near-optimal partition of antennas using statistical CSI. Then a lower bound and its approximation have been derived for the average mutual information. The lower bound plus a constant offers a very accurate approximation to the average mutual information, and the computational complexity of the lower bound and its approximation are a few orders of magnitudes less than that of the average mutual information. Furthermore, a manifold-based gradient ascent algorithm has been proposed to find optimal analog and digital precoders via maximizing the lower bound approximation of the average mutual information. Several numerical results have also been provided to show that our proposed algorithm outperforms existing hybrid precoding algorithms.

APPENDIX

Proof of Proposition 1: We start with the necessary condition, i.e., if $\mathbf{F}$ is a feasible point of problem (28), then $\mathbf{F}$ satisfies (29). Since $|\mathbf{F}_{ij}| = \mathbf{1}_{\mathcal{S}_j}(i)$, we have $|\mathbf{F}_{ij}| \in \{0, 1\}$. In addition, the following equations

$$\bigcup_{j=1}^{N_{\mathrm{rf}}} \mathcal{S}_j = \{1, 2, \ldots, N_{\mathrm{t}}\} \tag{68}$$

$$\mathcal{S}_j \cap \mathcal{S}_k = \emptyset, \quad \forall j \neq k \tag{69}$$

implies $\sum_{i=1}^{N_{\mathrm{t}}} \|\mathbf{F}_{i\bullet}\|_0 = N_{\mathrm{t}}$ and $\|\mathbf{F}_{i\bullet}\|_0 \leq 1$, respectively. Therefore, we have $\|\mathbf{F}_{i\bullet}\|_0 = 1$. This completes the first part of the proof.

Next, we prove the sufficient condition, i.e., if $\mathbf{F}$ satisfies (29), then $\mathbf{F}$ is a feasible point of problem (28). Let $\mathcal{S}_j$ denotes positions of nonzero entries in the jth column of $\mathbf{F}$. Since $|\mathbf{F}_{ij}| \in \{0, 1\}$, we can express $\mathbf{F}$ as

$$|\mathbf{F}_{ij}| = \begin{cases} 1 & \text{if } i \in \mathcal{S}_j \\ 0 & \text{otherwise.} \end{cases} \tag{70}$$

In addition, since $\|\mathbf{F}_{i\bullet}\|_0 = 1$ for $i \in \{1, 2, \ldots, n_{\mathrm{T}}\}$, $\{\mathcal{S}_j\}$ must satisfy

$$\bigcup_{j=1}^{N_{\mathrm{rf}}} \mathcal{S}_j = \{1, 2, \ldots, N_{\mathrm{t}}\} \tag{71}$$

$$\mathcal{S}_j \cap \mathcal{S}_\ell = \emptyset, \quad \forall j \neq \ell. \tag{72}$$

Therefore, $\{\mathcal{S}_j\}_{j=1}^{n_{\mathrm{RF}}} \in \mathcal{S}$ and this completes the proof. ∎

Proof of Proposition 2: Note that $\log(x)$ is a concave function for $x > 0$. Using Jensen's inequality, the average mutual information with finite-alphabet inputs $E_{\mathbf{H}} \mathcal{I}(\mathbf{x}; \mathbf{y} | \mathbf{H})$ can be lower bounded by

$$E_{\mathbf{H}} \mathcal{I}(\mathbf{x}; \mathbf{y} | \mathbf{H})$$
$$\geq \log K - N_{\mathrm{r}} \left(\frac{1}{\ln 2} - 1 \right)$$
$$- \frac{1}{K} \sum_{m=1}^{K} \log \sum_{k=1}^{K} E_{\mathbf{H}, \mathbf{n}} \exp \left(-\frac{\|\mathbf{e}_{mk} + \mathbf{n}\|^2}{\sigma^2} \right) \tag{73}$$

where $\mathbf{e}_{mk} = \mathbf{H} \bar{\mathbf{F}} \bar{\mathbf{B}} (\mathbf{x}_m - \mathbf{x}_k)$.

Since $\mathbf{n}$ is the i.i.d. complex Gaussian noise, the expectation over $\mathbf{n}$ in (73) can be calculated as

$$E_{\mathbf{n}} \exp \left(-\frac{\|\mathbf{e}_{mk} + \mathbf{n}\|^2}{\sigma^2} \right)$$
$$= \frac{1}{(\pi\sigma^2)^{N_{\mathrm{r}}}} \int_{\mathbf{n}} \exp \left(-\frac{\|\mathbf{e}_{mk} + \mathbf{n}\|^2 + \|\mathbf{n}\|^2}{\sigma^2} \right) \mathrm{d}\mathbf{n}$$
$$= \frac{1}{(\pi\sigma^2)^{N_{\mathrm{r}}}} \int_{\mathbf{n}} \prod_{i=1}^{N_{\mathrm{r}}} \exp \left(-\frac{|e_{mk,i} + n_i|^2 + |n_i|^2}{\sigma^2} \right) \mathrm{d}\mathbf{n}$$
$$= \prod_{i=1}^{N_{\mathrm{r}}} \frac{1}{\pi\sigma^2} \int_{n_i} \exp \left(-\frac{|e_{mk,i} + n_i|^2 + |n_i|^2}{\sigma^2} \right) \mathrm{d}n_i \tag{74}$$

where $e_{mk,i}$ and n_i are the ith element of $\mathbf{e}_{mk}$ and $\mathbf{n}$, respectively. Applying the integrals of exponential function and extending it to the complex-valued case, equation (74) is rewritten as

$$E_{\mathbf{n}} \exp \left(-\frac{\|\mathbf{e}_{mk} + \mathbf{n}\|^2}{\sigma^2} \right) = \prod_{i=1}^{N_{\mathrm{r}}} \frac{1}{2} \exp \left(-\frac{|e_{mk,i}|^2}{2\sigma^2} \right)$$
$$= \frac{1}{2^{N_{\mathrm{r}}}} \exp \left(-\frac{\mathbf{e}_{mk}^H \mathbf{e}_{mk}}{2\sigma^2} \right). \tag{75}$$

Then we insert $\mathbf{e}_{mk} = \left(\frac{N_{\mathrm{r}} N_{\mathrm{t}}}{L} \right)^{\frac{1}{2}} \mathbf{A}_{\mathrm{r}} \mathbf{\Gamma} \mathbf{A}_{\mathrm{t}}^H \bar{\mathbf{F}} \bar{\mathbf{B}} (\mathbf{x}_m - \mathbf{x}_k)$ into equation (75), and it yields

$$\exp \left(-\frac{\mathbf{e}_{mk}^H \mathbf{e}_{mk}}{2\sigma^2} \right) = \exp \left[-\mathrm{tr} \left(\mathbf{\Gamma} \mathbf{W}_{mk} \mathbf{\Gamma}^H \mathbf{A}_{\mathrm{r}}^H \mathbf{A}_{\mathrm{r}} \right) \right] \tag{76}$$

where $\mathbf{W}_{mk} = \frac{N_{\mathrm{r}} N_{\mathrm{t}}}{2\sigma^2 L} \mathbf{A}_{\mathrm{t}}^H \bar{\mathbf{F}} \bar{\mathbf{B}} (\mathbf{x}_m - \mathbf{x}_k)(\mathbf{x}_m - \mathbf{x}_k)^H \bar{\mathbf{B}}^H \bar{\mathbf{F}}^H \mathbf{A}_{\mathrm{t}}$. Since $\mathbf{\Gamma}$ is a diagonal matrix, we have

$$\mathrm{tr} \left(\mathbf{\Gamma} \mathbf{W}_{mk} \mathbf{\Gamma}^H \mathbf{A}_{\mathrm{r}}^H \mathbf{A}_{\mathrm{r}} \right) = \mathbf{d}^H \left[\left(\mathbf{A}_{\mathrm{r}}^H \mathbf{A}_{\mathrm{r}} \right)^T \circ \mathbf{W}_{mk} \right] \mathbf{d} \tag{77}$$

where $\mathbf{d}$ is the diagonal entries of $\mathbf{\Gamma}^*$. Since the diagonal entries of $\mathbf{\Gamma}$ are i.i.d. complex Gaussian distributed, $\mathbf{d}$ is an i.i.d. complex Gaussian vector. Then the expectation of equation (76) over $\mathbf{\Gamma}$ can be expressed as

$$E_{\mathbf{\Gamma}} \exp \left[-\mathrm{tr} \left(\mathbf{\Gamma} \mathbf{W}_{mk} \mathbf{\Gamma}^H \mathbf{A}_{\mathrm{r}}^H \mathbf{A}_{\mathrm{r}} \right) \right]$$
$$= E_{\mathbf{d}} \exp \left[-\mathbf{d}^H \left(\mathbf{A}_{\mathrm{r}}^H \mathbf{A}_{\mathrm{r}} \right)^T \circ \mathbf{W}_{mk} \mathbf{d} \right]$$
$$= \frac{1}{\pi^L} \int_{\mathbf{d}} \exp \left[-\mathbf{d}^H \left(\mathbf{I} + \left(\mathbf{A}_{\mathrm{r}}^H \mathbf{A}_{\mathrm{r}} \right)^T \circ \mathbf{W}_{mk} \right) \mathbf{d} \right] \mathrm{d}\mathbf{d}$$
$$= \det \left[\mathbf{I} + \left(\mathbf{A}_{\mathrm{r}}^H \mathbf{A}_{\mathrm{r}} \right)^T \circ \mathbf{W}_{mk} \right]^{-1}. \tag{78}$$

The combination of (73), (75), (76) and (78) yields the lower bound. This completes the proof. ∎

Proof of Proposition 3: Since $\lim_{N_{\mathrm{r}} \to \infty} \mathbf{A}_{\mathrm{r}}^H \mathbf{A}_{\mathrm{r}} = \mathbf{I}$, we can replace $\mathbf{A}_{\mathrm{r}}^H \mathbf{A}_{\mathrm{r}}$ by $\mathbf{I}$. Then $L(\bar{\mathbf{F}}, \bar{\mathbf{B}})$ can be approximated as

$$L(\bar{\mathbf{F}}, \bar{\mathbf{B}}) \approx \log K - N_{\mathrm{r}} \left(\frac{1}{\ln 2} - 1 \right)$$
$$- \frac{1}{K} \sum_{m=1}^{K} \log \sum_{k=1}^{K} \det \left[\mathbf{I} + \mathbf{I} \circ \mathbf{W}_{mk} \right]^{-1}. \tag{79}$$

Note that $\mathbf{I} + \mathbf{I} \circ \mathbf{W}_{mk}$ is a diagonal matrix, with the ℓth diagonal element being

$$\left[\mathbf{I} + \mathbf{I} \circ \mathbf{W}_{mk} \right]_{\ell\ell} = 1 + \frac{N_{\mathrm{r}} N_{\mathrm{t}}}{2\sigma^2 L} |\beta_{mk\ell}|^2 \tag{80}$$

where $\beta_{mk\ell} = \mathbf{a}(\theta_{\mathrm{t},\ell})^H \bar{\mathbf{F}}\bar{\mathbf{B}}(\mathbf{x}_m - \mathbf{x}_k)$. Combining (79) and (80) yields the lower bound approximation. This completes the proof. ∎

Proof of Proposition 4: We first rewrite $L_A(\bar{\mathbf{F}}, \bar{\mathbf{B}})$ as

$$L_A(\bar{\mathbf{F}}, \bar{\mathbf{B}}) = \log K - N_{\mathrm{r}}\left(\frac{1}{\ln 2} - 1\right) - \frac{1}{K}\sum_{m=1}^{K}$$
$$\times \log \sum_{k=1}^{K}\exp\left[-\sum_{\ell=1}^{L}\ln\left(1 + \frac{N_{\mathrm{r}}N_{\mathrm{t}}}{2\sigma^2 L}|\beta_{mk\ell}|^2\right)\right]. \tag{81}$$

Using the chain rule in differentiation, the differential of $\bar{\mathcal{I}}_A(\bar{\mathbf{F}}, \bar{\mathbf{B}})$ with respect to $\bar{\mathbf{P}} = \bar{\mathbf{F}}\bar{\mathbf{B}}$ is

$$\mathrm{d}L_A = \mathrm{tr}\left(\mathrm{d}\bar{\mathbf{P}}^H\mathbf{G} + \mathbf{G}^H\mathrm{d}\bar{\mathbf{P}}\right) \tag{82}$$

where $\mathbf{G} = \sum_{\ell=1}^{L}\mathbf{a}(\theta_{\mathrm{t},\ell})\mathbf{a}(\theta_{\mathrm{t},\ell})^H\bar{\mathbf{F}}\bar{\mathbf{B}}\mathbf{E}_\ell$. Inserting $\mathrm{d}\bar{\mathbf{P}} = \bar{\mathbf{F}}\mathrm{d}\bar{\mathbf{B}}$ into equation (82), we obtain

$$\mathrm{d}L_A = \left(\mathrm{d}\bar{\mathbf{B}}^H\bar{\mathbf{F}}^H\mathbf{G} + \mathbf{G}^H\bar{\mathbf{F}}\mathrm{d}\bar{\mathbf{B}}\right). \tag{83}$$

Since $R(\boldsymbol{\Phi}, \bar{\mathbf{B}}) \triangleq L_A(\bar{\mathbf{F}}(\boldsymbol{\Phi}), \bar{\mathbf{B}})$, the gradient of $R(\boldsymbol{\Phi}, \bar{\mathbf{B}})$ with respect to $\bar{\mathbf{B}}$ is

$$\nabla_{\bar{\mathbf{B}}}R(\boldsymbol{\Phi}, \bar{\mathbf{B}}) = \bar{\mathbf{F}}^H\mathbf{G}. \tag{84}$$

Similarly, inserting $\mathrm{d}\bar{\mathbf{P}} = \mathrm{d}\bar{\mathbf{F}}\bar{\mathbf{B}}$ into equation (82), we obtain

$$\mathrm{d}L_A = \mathrm{tr}\left(\mathrm{d}\bar{\mathbf{F}}^H\mathbf{G}\bar{\mathbf{B}}^H + \bar{\mathbf{B}}\mathbf{G}^H\mathrm{d}\bar{\mathbf{F}}\right). \tag{85}$$

In addition, since $\mathrm{d}\bar{\mathbf{F}}_{ij} = \jmath\bar{\mathbf{F}}_{ij}\mathrm{d}\boldsymbol{\Phi}_{ij}$, the differential of $\bar{\mathbf{F}}$ is given by

$$\mathrm{d}\bar{\mathbf{F}} = \jmath\mathrm{d}\boldsymbol{\Phi}\circ\bar{\mathbf{F}}. \tag{86}$$

Inserting $\mathrm{d}\bar{\mathbf{F}} = \jmath\mathrm{d}\boldsymbol{\Phi}\circ\bar{\mathbf{F}}$ into equation (85) and using the following two equations

$$R(\boldsymbol{\Phi}, \bar{\mathbf{B}}) \triangleq L_A(\bar{\mathbf{F}}(\boldsymbol{\Phi}), \bar{\mathbf{B}}) \tag{87}$$
$$\mathrm{tr}\left[(\mathbf{A}\circ\mathbf{B})\mathbf{C}\right] = \mathrm{tr}\left[\mathbf{A}(\mathbf{C}\circ\mathbf{B}^T)\right] \tag{88}$$

we conclude that

$$\nabla_{\boldsymbol{\Phi}}R(\boldsymbol{\Phi}, \bar{\mathbf{B}}) = 2\Im\left(\mathbf{G}\bar{\mathbf{B}}^H\circ\bar{\mathbf{F}}^*\right). \tag{89}$$

This completes the proof. ∎

REFERENCES

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

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

[3] X. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, Apr. 2016.

[4] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.

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

[6] R. Rajashekar and L. Hanzo, "Hybrid beamforming in mm-Wave MIMO systems having a finite input alphabet," *IEEE Trans. Commun.*, vol. 64, no. 8, pp. 3337–3349, Aug. 2016.

[7] R. Lopez-Valcarce, N. González-Prelcic, C. Rusu, and R. W. Heath, Jr., "Hybrid precoders and combiners for mmwave MIMO systems with per-antenna power constraints," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–6.

[8] R. Rajashekar and L. Hanzo, "Iterative matrix decomposition aided block diagonalization for mm-wave multiuser MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1372–1384, Mar. 2017.

[9] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "Hybrid precoding for millimeter wave cellular systems with partial channel knowledge," in *Proc. Inf. Theory Appl. Workshop (ITA)*, Feb. 2013, pp. 1–5.

[10] Z. Li, S. Han, and A. F. Molisch, "Optimizing channel-statistics-based analog beamforming for millimeter-wave multi-user massive MIMO downlink," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4288–4303, Jul. 2017.

[11] 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.

[12] P. D. Karamalis, N. D. Skentos, and A. G. Kanatas, "Adaptive antenna subarray formation for MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 5, no. 11, pp. 2977–2982, Nov. 2006.

[13] S. Sanayei and A. Nosratinia, "Antenna selection in MIMO systems," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 68–73, Oct. 2004.

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

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

[16] D. Guo, S. Shamai Shitz, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

[17] A. Lozano, A. M. Tulino, and S. Verdú, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3033–3051, Jul. 2006.

[18] C. Xiao and Y. R. Zheng, "On the mutual information and power allocation for vector Gaussian channels with finite discrete inputs," in *Proc. IEEE Global Telecommun. Conf.*, Nov./Dec. 2008, pp. 1–5.

[19] F. Pèrez-Cruz, M. R. D. Rodrigues, and S. Verdú, "MIMO Gaussian channels with arbitrary inputs: Optimal precoding and power allocation," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1070–1084, Mar. 2010.

[20] C. Xiao, Y. R. Zheng, and Z. Ding, "Globally optimal linear precoders for finite alphabet signals over complex vector Gaussian channels," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3301–3314, Jul. 2011.

[21] W. Zeng, C. Xiao, and J. Lu, "A low-complexity design of linear precoding for MIMO channels with finite-alphabet inputs," *IEEE Wireless Commun. Lett.*, vol. 1, no. 1, pp. 38–41, Feb. 2012.

[22] Y. Wu, C.-K. Wen, C. Xiao, X. Gao, and R. Schober, "Linear precoding for the MIMO multiple access channel with finite alphabet inputs and statistical CSI," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 983–997, Feb. 2015.

[23] Y. Wu, D. W. K. Ng, C. W. Wen, R. Schober, and A. Lozano, "Low-complexity MIMO precoding for finite-alphabet signals," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4571–4584, Jul. 2017.

[24] Y. Wu, J.-B. Wang, J. Wang, R. Schober, and C. Xiao, "Secure transmission with large numbers of antennas and finite alphabet inputs," *IEEE Trans. Commun.*, vol. 65, no. 8, pp. 3614–3628, Aug. 2017.

[25] J. Jin, C. Xiao, M. Tao, and W. Chen, "Linear precoding for fading cognitive multiple-access wiretap channel with finite-alphabet inputs," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3059–3070, Apr. 2017.

[26] J. Jin, Y. R. Zheng, W. Chen, and C. Xiao, "Generalized quadratic matrix programming: A unified framework for linear precoding with arbitrary input distributions," *IEEE Trans. Signal Process.*, vol. 65, no. 18, pp. 4887–4901, Sep. 2017.

[27] Y. Wu, C. Xiao, Z. Ding, X. Gao, and S. Jin, "A survey on MIMO transmission with finite input signals: Technical challenges, advances, and future trends," *Proc. IEEE*, vol. 106, no. 10, pp. 1779–1833, Oct. 2018.

[28] J. Jin *et al.*, "Hybrid precoding for millimeter wave MIMO systems: A matrix factorization approach," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 3327–3339, May 2018.

[29] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[30] X. Gao, L. Dai, Y. Zhang, T. Xie, X. Dai, and Z. Wang, "Fast channel tracking for terahertz beamspace massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 5689–5696, Jul. 2017.

[31] D. E. Knuth *et al.*, *Concrete Mathematics*. Reading, MA, USA: Adison Wesley, 1989.

[32] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1994.

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

[34] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2349–2360, Sep. 2005.

[35] P. V. Amadori and C. Masouros, "Low RF-complexity millimeter-wave beamspace-MIMO systems by beam selection," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2212–2223, Jun. 2015.

[36] C. A. Balanis, *Antenna Theory: Analysis and Design*. Hoboken, NJ, USA.: Wiley, 2012.

[37] T. S. Rappaport, J. N. Murdock, and F. Gutierrez, Jr., "State of the art in 60-GHz integrated circuits and systems for wireless communications," *Proc. IEEE*, vol. 99, no. 8, pp. 1390–1436, Aug. 2011.

Juening Jin received the B.S. degree in electronic engineering from Southeast University, Nanjing, China, in 2013, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2018.

From 2015 to 2017, he was a Research Scholar with the Missouri University of Science and Technology, Rolla, MO, USA. Since 2018, he has been with Huawei Technology, Shanghai, China, where he is currently a Senior System Engineer. His research interests include wireless communications and signal processing.
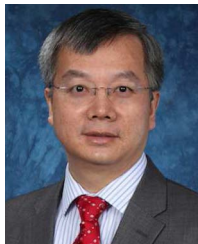
Chengshan Xiao (M'99–SM'02–F'10) received the B.Sc. degree in electronic engineering from the University of Electronic Science and Technology of China, in 1987, the M.Sc. degree in electronic engineering from Tsinghua University, in 1989, and the Ph.D. degree in electrical engineering from The University of Sydney, in 1997.

He has served as the Program Director of the Division of Electrical, Communications and Cyber Systems, USA National Science Foundation. He was a Senior Member of scientific staff at Nortel Networks, Ottawa, Canada. He was a Faculty Member at Tsinghua University, Beijing, China; the University of Alberta, Edmonton, Canada; the University of Missouri, Columbia, MO, USA, and the Missouri University of Science and Technology, Rolla, MO, USA. He has also held visiting professor positions in Germany and Hong Kong. He is currently the Chandler Weaver Professor, and also the Chair of the Department of Electrical and Computer Engineering, Lehigh University. His research interests include wireless communications, signal processing, and underwater acoustic communications. He holds several patents granted in USA, Canada, China, and Europe. His invented algorithms have been implemented into Nortel's base station radio products after successful technical field trials and network integration.

Dr. Xiao is a fellow of the Canadian Academy of Engineering. He has received several distinguished awards, including the 2014 Humboldt Research Award, the 2014 IEEE Communications Society Joseph LoCicero Award, the 2015 IEEE Wireless Communications Technical Committee Recognition Award, and the 2017 IEEE Communications Society Harold Sobol Award. He is the Awards Committee Chair and an elected Member-at-Large of the Board of Governors of the IEEE Communications Society. Previously, he has served on the IEEE Technical Activity Board Periodical Committee. He was an elected Member-at-Large of the Board of Governors, a member of the Fellow Evaluation Committee, the Director of Conference Publications, a Distinguished Lecturer of the IEEE Communications Society, and a Distinguished Lecturer of the IEEE Vehicular Technology. He also served as an Editor, Area Editor, and the Editor-in-Chief of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, an Associate Editor of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-I. He was the Technical Program Chair of the 2010 IEEE International Conference on Communications, Cape Town, South Africa, and the Technical Program Co-Chair of the 2017 IEEE Global Communications Conference, Singapore. He has served as the founding Chair of the IEEE Wireless Communications Technical Committee.

Wen Chen (M'03–SM'11) received the B.S. and M.S. degrees from Wuhan University, China, in 1990 and 1993, respectively, and the Ph.D. degree from the University of Electro-Communications, Tokyo, Japan, in 1999.

From 1999 to 2001, he was a JSPS fellow at the University of Electro-Communications. In 2001, he joined the University of Alberta, Canada, starting as a PIMS Post-Doctoral Fellow, and then as a Research Associate. Since 2006, he has been a Full Professor with the Department of Electronic Engineering, Shanghai Jiao Tong University, China, where he is also the Director of the Institute for Signal Processing and Systems. He has published 94 papers in IEEE journals and over 120 papers in IEEE conferences. His interests cover multiple access, coded cooperation, green heterogeneous networks, and vehicular communications.

Dr. Chen received the Ariyama Memorial Research Award, in 1997; and the JSPS Fellowship and PIMS Fellowship, in 1999 and 2001, respectively. He received the honors of "New Century Excellent Scholar in China" in 2006, and the "Pujiang Excellent Scholar in Shanghai" in 2007. He received the Best Service Award of the China Institute of Electronics, Shanghai Master Thesis Supervision Award, and the Best Paper Award of the Chinese Information Theory Society in 2013, the Best Presentation Paper Awards of the Chinese Information Theory Society in 2014, the Innovate 5G Competition Award and the IEEE WCSP best paper award in 2015, the IEEE APCC Keynote Speaker Appreciation in 2016, the IEEE VTS Chapter-of-the-Year Award and the Excellent PhD Thesis Supervision Award in 2018, and the IEEE ComSoc Distinguished Lecturer in 2019. He is the Chair of the IEEE VTS Shanghai Chapter. He is the TPC Chair and General Chair for many IEEE sponsored conferences. He is the Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON COMMUNICATIONS, and the IEEE ACCESS.

Yongpeng Wu (S'08–M'13–SM'17) received the B.S. degree in telecommunication engineering from Wuhan University, Wuhan, China, in 2007, and the Ph.D. degree in communication and signal processing from the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China, in 2013.

He is currently a Tenure-Track Associate Professor with the Department of Electronic Engineering, Shanghai Jiao Tong University, China. Previously, he was a a Senior Research Fellow with the Institute for Communications Engineering, Technical University of Munich, Germany, and the Humboldt Research Fellow and the Senior Research Fellow with the Institute for Digital Communications, University Erlangen-Nürnberg, Germany. During his doctoral studies, he has conducted cooperative research at the Department of Electrical Engineering, Missouri University of Science and Technology, Rollo, MO, USA. His research interests include massive MIMO/MIMO systems, physical layer security, signal processing for wireless communications, and multivariate statistical theory.

Dr. Wu has been a TPC member of various conferences, including Globecom, ICC, VTC, and PIMRC. He was a recipient of the IEEE Student Travel Grants for the IEEE International Conference on Communications (ICC) 2010, the Alexander von Humboldt Fellowship in 2014, the Travel Grants for the IEEE Communication Theory Workshop 2016, the Excellent Doctoral Thesis Awards of the China Communications Society 2016, the Exemplary Editor Award of the IEEE Communication Letters 2017, and the Young Elite Scientist Sponsorship Program by CAST 2017. He was an Exemplary Reviewer of the IEEE TRANSACTIONS ON COMMUNICATIONS in 2015, 2016, and 2018, respectively. He was the Lead Guest Editor of the special issue Physical Layer Security for 5G Wireless Networks of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, and the Guest Editor of the special issue Safeguarding 5G-and-Beyond Networks with Physical Layer Security of the IEEE WIRELESS COMMUNICATIONS. He is currently an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS and the IEEE COMMUNICATIONS LETTERS.