IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 7, 2017, accepted January 7, 2018, date of publication January 15, 2018, date of current version March 9, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2793223

Hardware-Efficient Hybrid Precoding for Millimeter Wave Systems With Multi-Feed Reflectarrays

ZHENGYI ZHOU[1], NING GE[1], (Member, IEEE), ZHAOCHENG WANG[1], (Senior Member, IEEE), AND SHENG CHEN[2,3], (Fellow, IEEE)

[1]Tsinghua National Laboratory for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
[2]School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K.
[3]King Abdulaziz University, Jeddah 21589, Saudi Arabia

Corresponding author: Zhaocheng Wang (zcwang@mail.tsinghua.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 91638205, in part by Shenzhen Subject Arrangements under Grant JCYJ20160331184124954, in part by the Shenzhen Peacock Plan under Grant 1108170036003286, in part by the Guangdong Science and Technology Planning Project under Grant 2014B010120001, in part by the Shenzhen Fundamental Research Project under Grant JCYJ20150401112337177, and in part by the Shenzhen Visible Light Communication System Key Laboratory under Grant ZDSYS20140512114229398.

ABSTRACT Current hybrid precoding schemes for large-scale antenna array-aided millimeter wave (mmWave) wireless systems adopt either full-connected or sub-connected architectures based on phased array antennas (PAAs). The former achieves excellent precoding performance but imposes considerably high hardware complexity, while the latter is relatively simple but suffers from high performance loss. In this paper, we exploit the new technology of reflectarray antennas, which among other advantages offers significant hardware efficiency gain, and propose a novel hybrid precoding architecture based on multi-feed reflectarray antennas (MRAs) for mmWave wireless systems, which is capable of achieving considerable hardware efficiency gain even over the low-hardware-complexity sub-connected architecture with PAAs. The mathematical model is provided for this novel hybrid precoding architecture, and an efficient alternating descent algorithm is developed to jointly design the analog and digital precoders for the proposed hybrid precoding scheme. Numerical results obtained demonstrate that our proposed precoding scheme with MRAs achieves much better precoding performance than its sub-connected counterpart with PAAs.

INDEX TERMS Millimeter wave, hybrid precoding, phased array antennas, reflectarray antennas, hardware complexity, precoding performance, mutual information.

I. INTRODUCTION

Wireless data traffic is projected to increase by 1000 fold in 2020 [1]. This exponentially increasing demand for higher data rate services creates unprecedented challenges for the next generation outdoor wireless communication systems. In order to meet the demand of enormous capacity increase, higher segments of the frequency spectrum must be exploited [2]. Millimeter wave (mmWave) with abundant frequency bandwidth offers a promising technology to address the shortage of spectrum resources in the future wireless networks [3], [4]. In order to effectively utilize the spectrum resources of the mmWave band, large-scale antenna arrays with precoding at transmitter and combining at receiver must be leveraged to combat severe pathloss experienced by mmWave signals. Fortunately, compared to a current typical wireless carrier, a typical mmWave carrier has ten-fold shorter wavelength, which enables an mmWave antenna array to pack proportionally more antenna elements into the same aperture size. This makes it practical to implement large-scale antenna array at the base station (BS) of mmWave wireless system. With large-scale antenna array, the BS is capable of providing extremely narrow beams with very high beamforming gains for surmounting severe mmWave pathloss so as to extend coverage at longer ranges and to establish reliable links [5], [6].

Hence a typical mmWave system employs a very large number of transmit antennas and a much smaller number of receive antennas to support a limited number of independent

data streams or users, where the number of independent data streams is generally much smaller than the number of receive antennas. The optimal joint precoding and combining design, commonly referred to as joint transceiver design in the literature [7], [8], is clearly intractable for such a large-scale antenna array based system. In practice, therefore, the design of precoding and the design of combining are carried out separately. Because the size of transmitter antenna array is much larger than the size of receiver antenna array, the design of precoding is much more challenging than the design of combining. In particular, the cost of deploying powerful full-digital precoding becomes too high. More specifically, in full-digital precoding architecture, each transmit antenna element requires an exclusive radio frequency (RF) chain. As a result, a full-digital precoding based mmWave wireless system imposes too high hardware requirements and consumes too much power, and it is uneconomical and impractical to implement [9]–[11].

Consequently, hybrid precoding for mmWave systems, which trades off achievable precoding performance with implementation cost, attracts much attention [12], [13]. With hybrid precoding, the signal processing operations are divided between the digital and analog domains. Specifically, a low-dimensional digital precoding is implemented at the baseband, while the analog part is completed by applying high-dimensional RF precoding via analog phase shifters, so that a large number of antenna elements can be driven by only a limited number of RF chains, while capable of supporting a number of independent data streams transmitted on a shared time-frequency resource block [14], [15]. Presently, most of hybrid precoding schemes adopt either full-connected or sub-connected architectures [16], as illustrated in Fig. 1, both of which are implemented via phased array antennas (PAAs). More specifically, in the full-connected architecture of Fig. 1 (a), each RF chain is required to connect with all the antenna elements via microstrip lines, which imposes considerable challenge to implement the feeding network (FN) of large-scale antenna array in mmWave band. With hundreds of microstrip lines and mmWave RF components as well as the complicated FN structure, the full-connected architecture with PAAs (FC-PAAs) of Fig. 1 (a) is difficult to realize, which limits its practical applicability [16]–[18], despite of its excellent precoding performance. By contrast, in the sub-connected architecture with PAAs (SC-PAAs) of Fig. 1 (b), each RF chain connects only to a subset of antennas. Consequently, the SC-PAAs requires much less hardware components but introduces high precoding performance loss, compared to the FC-PAAs. The RF hardware requirements of the both FC-PAAs and SC-PAAs are compared in Table 1.

Recently, the technology of multi-beam reflectarray antennas has been developed and has matured [19]–[32]. In comparison with other types of antenna arrays, such as PAAs, the advantages of reflectarray antennas include high gain, low profile, low mass, easy to fabricate and easy for circuitry integration [21]. A type of single-feed reflectarray

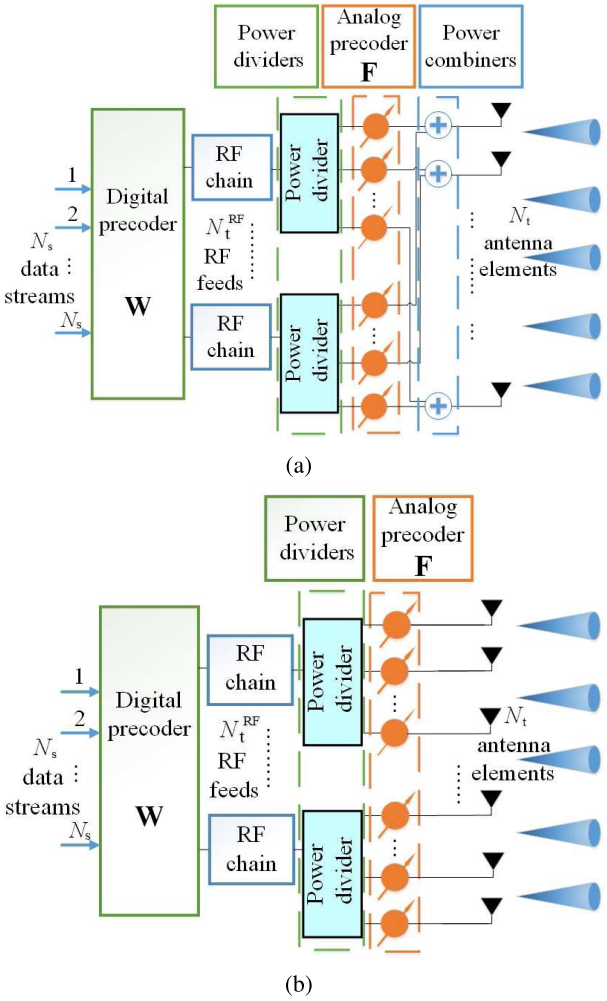

FIGURE 1. Conventional hybrid precoding architectures with phased array antennas: (a) full-connected model, and (b) sub-connected model, where the transmitter employs N_t antennas with N_t^{RF} ($\leq N_t$) RF chains to support N_s ($\leq N_t^{RF}$) independent data streams, and typically $N_t^{RF} \ll N_t$.

antennas (SRAs) was successfully designed and fabricated using particle swarm optimization [24]. Multi-feed reflectarray antennas (MRAs) can be implemented with low-complexity fabrication based on SRAs [25], [26]. Although conventional reflectarray antennas suffer from the major drawback of limited bandwidth in phase shifters, novel techniques such as nonresonant conductor (NRC) cell and metallic waveguide have been developed for bandwidth improvement so as to facilitate the applications of reflectarray antennas [28], [29]. Currently, such low-hardware-complexity reflectarray antennas have been widely used in radar detection, imaging systems, and satellite communications, and they can be further utilized in 5G mmWave wireless communications with proper design [29]–[32].

As is illustrated in Fig. 2 (a), MRAs leverage multiple RF feeds to transfer RF signals via irradiation instead of microstrip lines. Analog precoding is performed by phase shifters, one affiliated to an antenna element. This means that the number of phase shifters required is equal to the

TABLE 1. Radio frequency hardware complexity comparisons.

	Number of power dividers	Number of phase shifters	Number of combiners	RF signal transfer via
FC-PAAs	N_t^{RF}	$N_t \times N_t^{RF}$	N_t	Microstrip line
SC-PAAs	N_t^{RF}	N_t	0	Microstrip line
MRAs	0	N_t	0	Irradiation

N_t and N_t^{RF} are the numbers of antennas and RF chains, respectively.

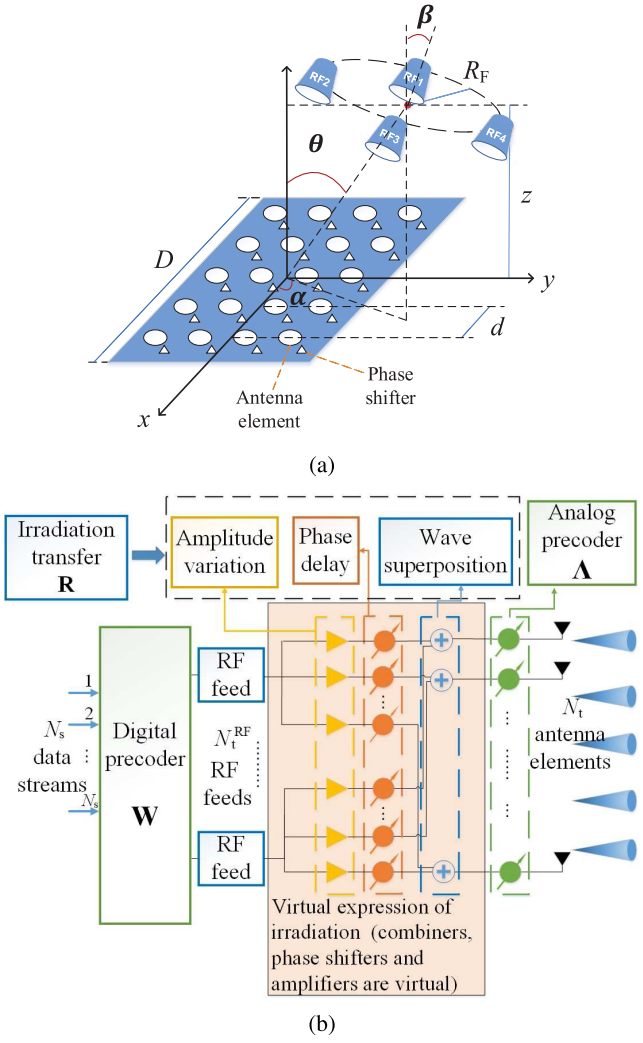

FIGURE 2. Hybrid precoding architecture with multi-feed reflectarray antennas: (a) architecture of multi-feed reflectarray antennas, and (b) transmitter diagram where the transmitter employs N_t antennas with N_t^{RF} ($\leq N_t$) RF feeds to support N_s ($\leq N_t^{RF}$) independent data streams, and typically $N_t^{RF} \ll N_t$.

number of antennas but no RF power divider or combiner is needed. Therefore, hybrid precoding with MRAs is capable of achieving considerable hardware efficiency gain, even over the low-hardware-complexity SC-PAAs, as can be seen from Table 1. However, applying hybrid precoding with MRAs to mmWave wireless systems has not been carried out in the existing literature and, moreover, it is unknown whether hybrid precoding with MRAs is superior or inferior to hybrid precoding with SC-PAAs, in terms of achievable precoding performance.

Therefore, we focus on low-hardware-complexity hybrid precoding with MRAs for mmWave systems in this paper. Note that since the design of precoding is separated from the design of combining, a precoder at the transmitter can operate with any properly designed combiner at the receiver, including the full-digital combiner and other hybrid combiners. Our main contributions are summarized as follows.

- We propose a novel hybrid precoding architecture based on MRAs, which removes the requirements of a large number of microstrip lines and other RF hardware components. Hence, our new hybrid precoding architecture achieves substantial hardware efficiency gain over the existing hybrid precoding architectures based on PAAs.
- We derive the mathematical model of the proposed hybrid precoding architecture, including the irradiation transfer process of MRAs and the analog precoding by the phase shifters affiliated to antenna elements. Then we formulate the design of the optimal digital and analog precoders that maximizes the mutual information (MI) subject to the analog precoder constraint.
- Because it is challenging to tackle directly this constrained joint precoding design optimization, we reformulate the joint precoding design as a constrained matrix recovery problem. We prove that due to its guaranteed asymptotic sufficiency, the solution of this reformulated constrained optimization problem is very close to the optimal solution of the original constrained joint precoding design problem.
- We show that the reformulated approximately equivalent constrained joint precoding design can be decomposed into the two analytically solvable sub-problems of analog and digital precoders, respectively. This enables us to develop an alternating descent algorithm to iteratively find a suboptimal solution to the joint design of the analog and digital precoders, with the guaranteed convergence.
- We evaluate the performance of the proposed MRAs-based hybrid precoding architecture in comparison with its conventional counterparts, the hybrid precoding schemes based on FC-PAAs and SC-PAAs, using an extensive simulation study.

Our investigating results demonstrate that the proposed hybrid precoding scheme based on MRAs, without the need of any RF power divider, power combiner and microstrip line, provides substantial hardware efficiency gain over the hybrid precoding scheme based on FC-PAAs, while only imposing

an acceptable precoding performance loss. More importantly, the proposed hybrid precoding scheme based on MRAs not only attains considerably better precoding performance than the hybrid precoding scheme based on SC-PAAs with the identical number of RF phase shifters but also offers a lower hardware complexity than the latter.

We adopt the following notations throughout this paper. The boldfaced upper-case and lower-case letters denote matrices and vectors, respectively, while calligraphic upper-case letters stands for sets. $\mathbb{R}^n$ denotes the n-dimensional real space and $\mathbb{C}^n$ is the n-dimensional complex space, while $\mathbb{C}^{m\times n}$ stands for the space of all the $m \times n$ complex matrices. The transpose, conjugate transpose and inverse operators are denoted by $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\mathrm{H}}$ and $(\cdot)^{-1}$, respectively, while $\mathrm{tr}(\cdot)$ and $\det(\cdot)$ denotes the matrix trace and determinant operators, respectively. The Frobenius norm of matrix $\boldsymbol{A}$ is denoted by $\|\boldsymbol{A}\|_F$, and $\boldsymbol{A}_{i,j}$ is the (i,j)th entry of $\boldsymbol{A}$, while $\|\boldsymbol{a}\|$ and $|a|$ denote the 2-norm of complex vector $\boldsymbol{a}$ and the magnitude of complex number a, respectively. The diagonal matrix with the diagonal entries $\{a_1, \cdots, a_n\}$ is denoted by $\mathrm{diag}\{a_1, \cdots, a_n\}$, $\boldsymbol{I}_n$ is the $n \times n$ identity matrix, and $\boldsymbol{0}_n$ is the n-dimensional all-zero vector, while we use $\boldsymbol{0}$ to denote all-zero vector or matrix of appropriate dimension. A complex number $a \in \mathbb{C}$ is represented either by $a = \Re[a] + \mathbf{j}\Im[a]$ or by $a = |a|e^{\mathbf{j}\angle a}$. Lastly, $\mathsf{E}\{\cdot\}$ denotes the expectation operator.

II. SYSTEM MODEL FOR MRAs

It is well-known that employing large-scale antenna arrays is highly effective in mitigating the inter-user interference with the aid of only simple linear signal processing, and in the asymptotic case of infinite many transmit antennas, the user accesses to the system becomes effectively orthogonal [33]. Therefore, for notational simplicity and without loss of generality, we consider a single-user mmWave system with large-scale antenna arrays where the transmitter employs N_{t} antennas to forward the N_{s} data streams to the receiver equipped with N_{r} antennas. The transmitter is equipped with $N_{\mathrm{t}}^{\mathrm{RF}}$ RF chains with $N_{\mathrm{s}} \le N_{\mathrm{t}}^{\mathrm{RF}} \le N_{\mathrm{t}}$. Typically, N_{t} is very large, and N_{r} is much smaller, that is, $N_{\mathrm{t}}^{\mathrm{RF}} \ll N_{\mathrm{t}}$ and $N_{\mathrm{r}} \ll N_{\mathrm{t}}$. Furthermore, generally $N_{\mathrm{s}} \ll N_{\mathrm{r}}$.

The transmitter of the proposed hybrid precoding architecture based on MRAs is illustrated in Fig. 2 (b). Let the transmit symbol vector be $\boldsymbol{s} \in \mathbb{C}^{N_{\mathrm{s}}}$ with $\mathsf{E}\{\boldsymbol{s}\boldsymbol{s}^{\mathrm{H}}\} = \frac{1}{N_{\mathrm{s}}}\boldsymbol{I}_{N_{\mathrm{s}}}$. As usual, digital precoding is performed at baseband with the digital precoding matrix $\boldsymbol{W} \in \mathbb{C}^{N_{\mathrm{t}}^{\mathrm{RF}}\times N_{\mathrm{s}}}$. The $N_{\mathrm{t}}^{\mathrm{RF}}$ RF feeds convert the digitally precoded baseband signals via irradiation transfer, which introduces amplitude variation and phase delay, and this physical transfer process is depicted in the dashed block at the top middle part of Fig. 2 (b). For a convenient comparison with the hybrid precoding architectures based on PAAs of Fig. 1, the actual irradiation transfer process is equivalent to the virtual RF process with the virtual RF components expressed virtually in the middle block of Fig. 2 (b). Finally, analog precoding is completed in the RF domain by the phase shifters affiliated to the N_{t} antenna elements, which are characterized by the $N_{\mathrm{t}} \times N_{\mathrm{t}}$ diagonal matrix $\boldsymbol{\Lambda} = \mathrm{diag}\left\{e^{\mathrm{j}\nu_1}, e^{\mathrm{j}\nu_2}, \cdots, e^{\mathrm{j}\nu_{N_{\mathrm{t}}}}\right\}$.

The irradiation transfer process is determined by the configuration of the RF feeds. Note that even for mmWave systems, it is physically impractical to implement very large-scale antenna arrays in one-dimensional form. Therefore, as illustrated in Fig. 2 (a), we adopt the uniform planar array (UPA) with the antenna spacing $d = 0.5\lambda$, where λ is the carrier's wavelength, and the feeds are uniformly located on a circle with radius R_{F}, and the center has the spherical coordinates of $\left(\frac{z}{\cos(\theta)}, \theta, \alpha\right)$, where z is the vertical height away from the array plane, θ is the polar angle and α is the azimuth angle of the feed circle center, respectively. The elevation angle of the feed circle to the array plane is β, and the array size is denoted by $D = \min\{N_{\mathrm{h}}d, N_{\mathrm{v}}d\}$, where N_h and N_v denote the numbers of 'row' and 'column' antenna elements, respectively, and $N_{\mathrm{t}} = N_{\mathrm{h}}N_{\mathrm{v}}$. The coordinates of the kth RF feed $\left(x_k^{\mathrm{RF}}, y_k^{\mathrm{RF}}, z_k^{\mathrm{RF}}\right)$ are therefore given by

$$\begin{pmatrix} x_k^{\mathrm{RF}} \\ y_k^{\mathrm{RF}} \\ z_k^{\mathrm{RF}} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\beta & -\sin\beta \\ 0 & \sin\beta & \cos\beta \end{pmatrix} \begin{pmatrix} R_{\mathrm{F}}\cos\frac{2\pi(k-1)}{N_{\mathrm{t}}^{\mathrm{RF}}} \\ R_{\mathrm{F}}\sin\frac{2\pi(k-1)}{N_{\mathrm{t}}^{\mathrm{RF}}} \\ 0 \end{pmatrix} + \begin{pmatrix} z\tan\theta\cos\alpha \\ z\tan\theta\sin\alpha \\ z \end{pmatrix}, \quad (1)$$

while the antenna element at the mth row and nth column of the UPA has the coordinates

$$\left(x_{m,n}^{\mathrm{a}}, y_{m,n}^{\mathrm{a}}, z_{m,n}^{\mathrm{a}}\right) = \left(\left(m - 1 - \frac{N_{\mathrm{h}}-1}{2}\right)d, \left(n - 1 - \frac{N_{\mathrm{v}}-1}{2}\right)d, 0\right). \quad (2)$$

We can number the N_{t} antennas of the MRAs by the antenna index $1 \le i \le N_{\mathrm{t}}$, with

$$i = (m-1)N_{\mathrm{v}} + n, \quad 1 \le m \le N_{\mathrm{h}},\ 1 \le n \le N_{\mathrm{v}}. \quad (3)$$

According to the feed horn radiation pattern model [24], the kth stream of RF signal transferred to the ith antenna goes through a phase delay of $\angle R_{i,k}$ and undergoes an amplitude variation of $|R_{i,k}|$. The phase delay is determined by the optical path difference as

$$\angle R_{i,k} = 2\pi L_{i,k}/\lambda, \quad (4)$$

where $L_{i,k}$ is the distance between the kth RF feed and the ith antenna given by

$$L_{i,k} = \sqrt{(x_k^{\mathrm{RF}} - x_i^{\mathrm{a}})^2 + (z_k^{\mathrm{RF}} - z_i^{\mathrm{a}})^2 + (z_k^{\mathrm{RF}} - z_i^{\mathrm{a}})^2}, \quad (5)$$

and $\left(x_i^{\mathrm{a}}, y_i^{\mathrm{a}}, z_i^{\mathrm{a}}\right)$ are the coordinates of the ith antenna. The amplitude variation is determined as

$$|R_{i,k}| = \cos^q(\Omega_{i,k}), \quad (6)$$

where q is a system parameter and $\Omega_{i,k} = \arccos\left(z_k^{\mathrm{RF}}/L_{i,k}\right)$ is the elevation angle between the kth RF feed and the ith antenna element [24]. The irradiation transfer process

shown in Fig. 2 (b) is therefore specified by the transfer matrix $\boldsymbol{R} \in \mathbb{C}^{N_{\rm t} \times N_{\rm t}^{\rm RF}}$, whose ith-row and kth-column entry is $\boldsymbol{R}_{i,k} = R_{i,k} = |R_{i,k}| e^{\mathrm{j}\angle R_{i,k}}$. The full column rank $\boldsymbol{R}$ is normalized to $\|\boldsymbol{R}\|_F^2 = N_{\rm t}^{\rm RF}$. The transfer matrix $\boldsymbol{R}$ is fixed once the antenna array is fabricated. That is, the values of all its elements $R_{i,k}$ are determined once the spatial configuration of the RF feeds are specified. In practice, $\boldsymbol{R}$ can be measured after the antenna array is installed, and our numerical results indicate that hybrid precoding performance is insensitive to $\boldsymbol{R}$ as long as the RF feeds are configured within a reasonable spatial scope, as will shown later in the simulation study section.

The received signal vector $\boldsymbol{y} \in \mathbb{C}^{N_{\rm r}}$ at the receiver's $N_{\rm r}$ antennas can be expressed by [13]

$$\boldsymbol{y} = \sqrt{\rho} \boldsymbol{H} \boldsymbol{\Lambda} \boldsymbol{R} \boldsymbol{W} \boldsymbol{s} + \boldsymbol{\xi}, \tag{7}$$

where ρ is the average received power, and the noise vector $\boldsymbol{\xi} \in \mathbb{C}^{N_{\rm r}}$ follows a complex symmetric $N_{\rm r}$-dimensional Gaussian distribution with the zero mean vector $\boldsymbol{0}_{N_{\rm t}}$ and the covariance matrix $\boldsymbol{\Xi} \in \mathbb{C}^{N_{\rm r} \times N_{\rm r}}$, i.e., $\boldsymbol{\xi} \sim \mathcal{CN}\left(\boldsymbol{0}_{N_{\rm t}}, \boldsymbol{\Xi}\right)$, while the channel matrix $\boldsymbol{H} \in \mathbb{C}^{N_{\rm r} \times N_{\rm t}}$ satisfies $\mathsf{E}\left\{\|\boldsymbol{H}\|_F^2\right\} = N_{\rm r} N_{\rm t}$. The transmission power constraint is ensured by normalizing the digital precoder $\boldsymbol{W}$ such that $\|\boldsymbol{\Lambda R W}\|_F^2 = N_{\rm s}$, which is equivalent to $\|\boldsymbol{R W}\|_F^2 = N_{\rm s}$.

It is well-known that mmWave wireless propagation experiences inadequate spatial selectivity owing to high pathloss. Considering limited scattering features and high levels of antenna correlations, a clustered extended Saleh-Valenzuela model is established, which has the $N_{\rm c}$ clusters and each cluster has the $N_{\rm p}$ paths. Thus, $\boldsymbol{H}$ is given by [13]

$$\boldsymbol{H} = \gamma \sum_{k=1}^{N_{\rm c}} \sum_{i=1}^{N_{\rm p}} \alpha_{i,k} u_{\rm r}\left(\phi_{i,k}^{\rm r}, \varphi_{i,k}^{\rm r}\right) u_{\rm t}\left(\phi_{i,k}^{\rm t}, \varphi_{i,k}^{\rm t}\right) \times \boldsymbol{a}_{\rm r}\left(\phi_{i,k}^{\rm r}, \varphi_{i,k}^{\rm r}\right) \boldsymbol{a}_{\rm t}^{\rm H}\left(\phi_{i,k}^{\rm t}, \varphi_{i,k}^{\rm t}\right), \tag{8}$$

where $\alpha_{i,k} \sim \mathcal{CN}(0, P_{i,k})$ is the complex gain of the ith path from the kth cluster, and $\phi_{i,k}^{\rm r}$ ($\varphi_{i,k}^{\rm r}$) and $\phi_{i,k}^{\rm t}$ ($\varphi_{i,k}^{\rm t}$) are the azimuth (elevation) angles of arrival and departure of the ith propagation path contributed by the kth cluster, while $u_{\rm r}(\phi, \varphi)$ ($u_{\rm t}(\phi, \varphi)$) represent the receive (transmit) antenna element gains at the specific angles (ϕ, φ), and $\boldsymbol{a}_{\rm r}(\phi, \varphi) \in \mathbb{C}^{N_{\rm r} \times 1}$ ($\boldsymbol{a}_{\rm t}(\phi, \varphi) \in \mathbb{C}^{N_{\rm t} \times 1}$) are the normalized transmit (receive) antenna array response vectors determined by the geometric architectures of the transmit (receive) antenna arrays. For the UPA with size of $N_{\rm h} \times N_{\rm v}$, the array response vector $\boldsymbol{a}_{\rm UPA}(\phi, \varphi)$ can be expressed as

$$\boldsymbol{a}_{\rm UPA}(\phi, \varphi) = \frac{1}{\sqrt{N_t}} \Big[1 \cdots e^{\mathrm{j}\frac{2\pi}{\lambda} d(m \sin\phi \sin\varphi + n \cos\varphi)} \cdots \quad e^{\mathrm{j}\frac{2\pi}{\lambda} d((N_{\rm h}-1)\sin\phi \sin\varphi + (N_{\rm v}-1)\cos\varphi)}\Big]^{\rm T}, \tag{9}$$

where $0 \le m \le (N_{\rm h} - 1)$ and $0 \le n \le (N_{\rm v} - 1)$. Finally, γ in (8) is a normalizing scalar factor to ensure $\mathsf{E}\left\{\|\boldsymbol{H}\|_F^2\right\} = N_{\rm r} N_{\rm t}$.

Many methods can be used to estimate the channel state information (CSI) [12], [34]–[37], and in this paper, we assume that the accurate CSI is available at the transmitter. We point out that to address the CSI uncertainty at transmitter caused by CSI feedback delay and quantization errors, robust precoding techniques have to be applied [38]–[42], which is however beyond the scope of our paper.

III. HYBRID PRECODING WITH MRAs

The effective MI achieved by Gaussian signaling can be adopted as the optimization objective of transmitter hybrid precoding design [13], which is expressed as

$$I(\boldsymbol{\Lambda}, \boldsymbol{W}) = \log_2\left(\det\left(\boldsymbol{I}_{N_{\rm r}} + \frac{\rho}{\sigma^2 N_{\rm s}} \boldsymbol{H} \boldsymbol{\Lambda} \boldsymbol{R} \boldsymbol{W} \left(\boldsymbol{H} \boldsymbol{\Lambda} \boldsymbol{R} \boldsymbol{W}\right)^{\rm H}\right)\right), \tag{10}$$

where the channel noise vector is assumed to be white, i.e., $\boldsymbol{\Xi} = \sigma^2 \boldsymbol{I}_{N_{\rm r}}$.

A. OPTIMAL JOINT DESIGN OF DIGITAL AND ANALOG PRECODERS

Based on the MI metric (10), the transmitter hybrid precoding design can be posed as the following constrained optimization problem:

$$\begin{aligned} \left(\boldsymbol{\Lambda}_{\rm opt}, \boldsymbol{W}_{\rm opt}\right) = \arg \max_{\boldsymbol{\Lambda} \in \mathbb{C}^{N_{\rm t} \times N_{\rm t}}, \boldsymbol{W} \in \mathbb{C}^{N_{\rm t}^{\rm RF} \times N_{\rm s}}} & I(\boldsymbol{\Lambda}, \boldsymbol{W}), \\ \text{s.t. } \boldsymbol{\Lambda} \in \mathcal{D} \triangleq & \left\{\boldsymbol{\Lambda} = \mathrm{diag}\left\{e^{\mathrm{j}v_1}, \cdots, e^{\mathrm{j}v_{N_{\rm t}}}\right\}\right\}, \\ & \|\boldsymbol{R W}\|_F^2 = N_{\rm s}. \end{aligned} \tag{11}$$

The joint digital and analog precoders' optimization (11) is non-convex, owing to the constraint imposed on the analog precoder $\boldsymbol{\Lambda}$, which is challenging to solve directly even in a suboptimal way. Therefore, we propose to consider alternative joint digital and analog precoders optimization, whose optimal solution is at least very close to the optimal solution of the challenging optimization problem (11) and, moreover, is easier to tackle.

First, consider the normalized MI of the full precoder $\boldsymbol{X} \in \mathbb{C}^{N_{\rm t} \times N_{\rm s}}$ defined as

$$\bar{I}(\boldsymbol{X}) = \log_2\left(\det\left(\boldsymbol{I}_{N_{\rm r}} + \frac{1}{\mathrm{tr}\left(\boldsymbol{X}\boldsymbol{X}^{\rm H}\right)} \boldsymbol{H} \boldsymbol{X} \boldsymbol{X}^{\rm H} \boldsymbol{H}^{\rm H}\right)\right). \tag{12}$$

Then the optimal full digital precoding solution is the optimal unconstrained unitary precoder $\boldsymbol{X}_{\rm opt} \in \mathbb{C}^{N_{\rm t} \times N_{\rm s}}$ that maximizes $\bar{I}(\boldsymbol{X})$ [13], namely,

$$\boldsymbol{X}_{\rm opt} = \arg \max_{\boldsymbol{X} \in \mathbb{C}^{N_{\rm t} \times N_{\rm s}}} \bar{I}(\boldsymbol{X}). \tag{13}$$

As aforementioned, the full digital precoder requires $N_{\rm t}^{\rm RF} = N_{\rm t}$ and imposes too high hardware requirements. Consequently, it is impractical to implement for large-scale antenna array aided mmWave systems. However, given the CSI $\boldsymbol{H}$, the closed-form $\boldsymbol{X}_{\rm opt}$ can readily be obtained analytically from (13) [13]. Therefore, we can utilize the optimal full precoder solution $\boldsymbol{X}_{\rm opt}$ as the 'desired response' for our joint digital and analog precoders' design. More specifically, we can joint optimize the analog and digital precoders $\boldsymbol{\Lambda}$ and $\boldsymbol{W}$ so that $\boldsymbol{\Lambda R W}$ is as close as possible to $\boldsymbol{X}_{\rm opt}$. Of course

we have to consider the constraint imposed on the analog precoder. This leads to the following alternative constrained optimization for joint design of $\boldsymbol{\Lambda}$ and $\boldsymbol{W}$:

$$\begin{aligned}\left(\widetilde{\boldsymbol{\Lambda}}_{\text{opt}}, \widetilde{\boldsymbol{W}}_{\text{opt}}\right) = \arg \min_{\boldsymbol{\Lambda}\in\mathbb{C}^{N_\text{t}\times N_\text{t}}, \boldsymbol{W}\in\mathbb{C}^{N_\text{t}^\text{RF}\times N_\text{s}}} \|\boldsymbol{X}_{\text{opt}} - \boldsymbol{\Lambda}\boldsymbol{R}\boldsymbol{W}\|_F^2,\\ \text{s.t. } \boldsymbol{\Lambda}\in\mathcal{D},\\ \|\boldsymbol{R}\boldsymbol{W}\|_F^2 = N_\text{s}.\end{aligned} \tag{14}$$

We now prove that the optimal solution of the constrained optimization problem (14) is very closed to the optimal solution of the constrained optimization problem (11), that is, $\left(\widetilde{\boldsymbol{\Lambda}}_{\text{opt}}, \widetilde{\boldsymbol{W}}_{\text{opt}}\right)$ is at least a near-optimal solution to the original optimization problem (11). First, we introduce the following two lemmas.

Lemma 1: Given $\boldsymbol{A}, \boldsymbol{P} \in \mathbb{C}^{N\times K}$, define $\boldsymbol{D} = \boldsymbol{A}\boldsymbol{P}^\text{H} + \boldsymbol{P}\boldsymbol{A}^\text{H} \in \mathbb{C}^{N\times N}$. Then the trace $\text{tr}\left(\boldsymbol{D}^\text{H}\boldsymbol{D}\right)$ is bounded by four times of $\text{tr}\left(\boldsymbol{A}^\text{H}\boldsymbol{A}\right)\text{tr}\left(\boldsymbol{P}^\text{H}\boldsymbol{P}\right)$, i.e.,

$$\text{tr}\left(\boldsymbol{D}^\text{H}\boldsymbol{D}\right) \le 4\cdot\text{tr}\left(\boldsymbol{A}^\text{H}\boldsymbol{A}\right)\text{tr}\left(\boldsymbol{P}^\text{H}\boldsymbol{P}\right). \tag{15}$$

Proof: See Appendix A. ∎

Let the eigenvalues of a matrix $\boldsymbol{G}$ be arranged in the descending order. Further denote the ith largest eigenvalue of $\boldsymbol{G}$ as $\lambda_i(\boldsymbol{G})$ and the ith largest singular value of $\boldsymbol{G}$ as $\sigma_i(\boldsymbol{G})$.

Lemma 2: Given the complex matrices $\boldsymbol{H}$ and $\boldsymbol{D}$ with appropriate dimensions, the eigenvalue $\lambda_i\left(\boldsymbol{H}\boldsymbol{D}\boldsymbol{H}^\text{H}\right)$ is bounded, i.e., $\left|\lambda_i\left(\boldsymbol{H}\boldsymbol{D}\boldsymbol{H}^\text{H}\right)\right| \le \nu, \forall i$, where

$$\nu \triangleq \frac{1}{2} + \frac{1}{2}\text{tr}\left(\boldsymbol{D}^\text{H}\boldsymbol{D}\right)\left(\text{tr}\left(\boldsymbol{H}^\text{H}\boldsymbol{H}\right)\right)^2. \tag{16}$$

Proof: See Appendix B. ∎

Next we define the two metric spaces $\mathcal{B}\left(\mathbb{C}^{N_\text{t}\times N_\text{s}}, \|\cdot\|_F\right)$ and $\mathcal{B}(\mathbb{R}, |\cdot|)$ as the domain and range of the normalized MI $\bar{I}(\boldsymbol{X})$. Proposition 1 establishes the continuity of the MI mapping $\bar{I}(\cdot)$ in the F-normed domain based on Lemmas 1 and 2.

Proposition 1: The mutual information mapping

$$\begin{aligned}\bar{I}(\cdot) : \mathbb{C}^{N_\text{t}\times N_\text{s}} &\to \mathbb{R}\\ \boldsymbol{X} &\to \bar{I}(\boldsymbol{X})\end{aligned} \tag{17}$$

is continuous from $\mathcal{B}\left(\mathbb{C}^{N_\text{t}\times N_\text{s}}, \|\cdot\|_F\right)$ to $\mathcal{B}(\mathbb{R}, |\cdot|)$.

Proof: See Appendix C. ∎

Note that the normalized MI (12) and the original MI (10) are equivalent if we consider the full precoder as $\boldsymbol{X} = \boldsymbol{\Lambda}\boldsymbol{R}\boldsymbol{W}$ and it is normalized to $\text{tr}\left(\boldsymbol{X}\boldsymbol{X}^\text{H}\right) = \frac{\sigma^2 N_\text{s}}{\rho}$. Therefore, as a direct consequence of the continuity of $\bar{I}(\boldsymbol{X})$, and similar to the proofs of [15, eqs. (12) – (15)] based on Grassmann manifold, it can be confirmed that the optimal solution of the constrained optimization problem (14), $\left(\widetilde{\boldsymbol{\Lambda}}_{\text{opt}}, \widetilde{\boldsymbol{W}}_{\text{opt}}\right)$, is very close to the optimal solution of the original constrained optimization problem (11), $\left(\boldsymbol{\Lambda}_{\text{opt}}, \boldsymbol{W}_{\text{opt}}\right)$. Thus, instead of solving the problem (11), we can design the hybrid precoders by solving the problem (14). The trouble is that solving the constrained optimization problem (14) is also challenging because of the non-convex constraint $\boldsymbol{\Lambda} \in \mathcal{D}$, and it is difficult if not impossible to obtain the closed-form joint optimal design of the digital and analog precoders. However, the constrained optimization problem (14) can be tackled more easily in a suboptimal way, compared to the original problem (11).

B. SUBOPTIMAL DESIGN OF DIGITAL AND ANALOG PRECODERS

Since it is unrealistic to search for the closed-form optimal solution to the joint digital and analog precoders' design problem in the non-convex space $\mathcal{D} \times \mathbb{C}^{N_\text{t}^\text{RF}\times N_\text{s}}$, we adopt a two-stage alternating descent approach to iteratively design suboptimal hybrid precoders. Specifically, when the digital precoder $\boldsymbol{W}$ is fixed, the constrained optimization problem (14) is reduced to

$$\begin{aligned}\widehat{\boldsymbol{\Lambda}}_{\text{opt}} = \arg \min_{\boldsymbol{\Lambda}\in\mathbb{C}^{N_\text{t}\times N_\text{t}}} \|\boldsymbol{X}_{\text{opt}} - \boldsymbol{\Lambda}\boldsymbol{R}\boldsymbol{W}\|_F^2,\\ \text{s.t. } \boldsymbol{\Lambda}\in\mathcal{D},\end{aligned} \tag{18}$$

while given the analog precoder $\boldsymbol{\Lambda}$, the constrained optimization problem (14) becomes

$$\begin{aligned}\widehat{\boldsymbol{W}}_{\text{opt}} = \arg \min_{\boldsymbol{W}\in\mathbb{C}^{N_\text{t}^\text{RF}\times N_\text{s}}} \|\boldsymbol{X}_{\text{opt}} - \boldsymbol{\Lambda}\boldsymbol{R}\boldsymbol{W}\|_F^2,\\ \text{s.t. } \|\boldsymbol{R}\boldsymbol{W}\|_F^2 = N_\text{s}.\end{aligned} \tag{19}$$

Both the sub-problems are analytically solvable, and this enables us to achieve a suboptimal solution of the constrained optimization problem (14) iteratively.

For the first sub-problem, we can capture the diagonal structure and constant-norm property of $\boldsymbol{\Lambda}$ to reformulate it as an unconstrained real-valued vector optimization problem

$$\begin{aligned}\widehat{\boldsymbol{v}}^{\text{opt}} &= \arg \min_{\boldsymbol{v}\in\mathbb{R}^{N_\text{t}}} \left\|\boldsymbol{X}_{\text{opt}} - \text{diag}\left\{e^{\text{j}v_1}, \cdots, e^{\text{j}v_{N_\text{t}}}\right\}\boldsymbol{B}\right\|_F^2\\ &= \arg \min_{\boldsymbol{v}\in\mathbb{R}^{N_\text{t}}} \sum_{i=1}^{N_\text{t}} \|\boldsymbol{x}_i - e^{\text{j}v_i}\boldsymbol{b}_i\|^2,\end{aligned} \tag{20}$$

where $\boldsymbol{v} = \left[v_1\ v_2\cdots v_{N_\text{t}}\right]^\text{T}$ and $\boldsymbol{B} = \boldsymbol{R}\boldsymbol{W}$, while $\boldsymbol{x}_i^\text{T}$ is the ith row of $\boldsymbol{X}_{\text{opt}}$ and $\boldsymbol{b}_i^\text{T}$ is the ith row of $\boldsymbol{B}$ for $1 \le i \le N_\text{t}$. This vector optimization can be decoupled into the N_t scalar optimization problems:

$$\widehat{v}_i^{\text{opt}} = \arg \min_{v_i\in\mathbb{R}} \left\|\boldsymbol{x}_i - e^{\text{j}v_i}\boldsymbol{b}_i\right\|^2, \quad 1 \le i \le N_\text{t}. \tag{21}$$

Denote $\boldsymbol{x}_i^\text{H}\boldsymbol{b}_i = r_i e^{\text{j}\psi_i}$, where $r_i = \left|\boldsymbol{x}_i^\text{H}\boldsymbol{b}_i\right|$ and $\psi_i = \angle\boldsymbol{x}_i^\text{H}\boldsymbol{b}_i$ with $\psi_i \in [0,\ 2\pi)$. By simple algebraic manipulation, it can be seen that minimizing $\left\|\boldsymbol{x}_i - e^{\text{j}v_i}\boldsymbol{b}_i\right\|^2$ is equivalent to maximizing $\Re\left[e^{\text{j}v_i}\boldsymbol{x}_i^\text{H}\boldsymbol{b}_i\right]$. Moreover, we have

$$\Re\left[e^{\text{j}v_i}\boldsymbol{x}_i^\text{H}\boldsymbol{b}_i\right] = \Re\left[e^{\text{j}v_i} r_i e^{\text{j}\psi_i}\right] = r_i\cos(v_i + \psi_i) \le r_i. \tag{22}$$

In (22), the equality holds in the last inequality if and only if $\cos(v_i + \psi_i) = 1$, and this condition states that the optimal $e^{\text{j}v_i}$ is in the opposite direction of $\boldsymbol{x}_i^\text{H}\boldsymbol{b}_i$ on the complex plane. Hence, we can define the function $\mathcal{L}$ for the analog precoder in each iteration as

$$\begin{aligned}\mathcal{L}\left(\boldsymbol{W}; \boldsymbol{X}_{\text{opt}}, \boldsymbol{R}\right) = \text{diag}\left\{e^{\text{j}v_1}, e^{\text{j}v_1}, \cdots, e^{\text{j}v_{N_\text{t}}}\right\},\\ \text{with } e^{\text{j}v_i} = \frac{\boldsymbol{b}_i^\text{H}\boldsymbol{x}_i}{|\boldsymbol{b}_i^\text{H}\boldsymbol{x}_i|}, \quad 1 \le i \le N_\text{t}.\end{aligned} \tag{23}$$

For the second sub-problem, we can first convert it into an unconstrained least squares (LS) problem with the solution

$$\widetilde{\boldsymbol{W}} = \left(\boldsymbol{R}^{\mathrm{H}}\boldsymbol{R}\right)^{-1}\boldsymbol{R}^{\mathrm{H}}\boldsymbol{\Lambda}^{\mathrm{H}}\boldsymbol{X}_{\mathrm{opt}}, \tag{24}$$

and normalize the solution to met the total power requirement:

$$\widehat{\boldsymbol{W}}_{\mathrm{opt}} = \frac{\sqrt{N_{\mathrm{s}}}}{\|\boldsymbol{R}\widetilde{\boldsymbol{W}}\|_F}\widetilde{\boldsymbol{W}}. \tag{25}$$

Similarly, we can define the function $\mathcal{W}$ for the digital precoder in each iteration as

$$\mathcal{W}\left(\boldsymbol{\Lambda}; \boldsymbol{X}_{\mathrm{opt}}, \boldsymbol{R}\right) = \frac{\sqrt{N_{\mathrm{s}}}}{\|\boldsymbol{R}\widetilde{\boldsymbol{W}}\|_F}\widetilde{\boldsymbol{W}}. \tag{26}$$

The functions $\mathcal{L}$ and $\mathcal{W}$ are utilized to alternately update the iterative values of $\boldsymbol{\Lambda}$ and $\boldsymbol{W}$. Define the iterative decrement as

$$\delta_k = \|\boldsymbol{X}_{\mathrm{opt}} - \boldsymbol{\Lambda}^{(k-1)}\boldsymbol{R}\boldsymbol{W}^{(k-1)}\|_F - \|\boldsymbol{X}_{\mathrm{opt}} - \boldsymbol{\Lambda}^{(k)}\boldsymbol{R}\boldsymbol{W}^{(k)}\|_F. \tag{27}$$

Then $\delta_k < \epsilon$ serves as the stopping criterion for a suitably preset threshold ϵ. The proposed alternating decent method is listed in Algorithm 1, and the convergence properties of this proposed algorithm are detailed in Proposition 2.

Algorithm 1 Alternating Descent Method

Require: Unconstrained optimal full precoder $\boldsymbol{X}_{\mathrm{opt}}$, transfer matrix $\boldsymbol{R}$, termination threshold ϵ, and maximum number of iterations K

$k = 0$
$\delta_k = 2\epsilon$
$\boldsymbol{\Lambda}^{(k)} = \boldsymbol{I}_{N_{\mathrm{t}}}$
$\boldsymbol{W}^{(k)} = \mathcal{W}(\boldsymbol{\Lambda}^{(k)}; \boldsymbol{X}_{\mathrm{opt}}, \boldsymbol{R})$
for $\delta_k \geq \epsilon$ and $k < K$ **do**
  $k = k + 1$
  $t = \|\boldsymbol{X}_{\mathrm{opt}} - \boldsymbol{\Lambda}^{(k-1)}\boldsymbol{R}\boldsymbol{W}^{(k-1)}\|_F$
  $\boldsymbol{\Lambda}^{(k)} = \mathcal{L}(\boldsymbol{W}^{(k-1)}; \boldsymbol{X}_{\mathrm{opt}}, \boldsymbol{R})$
  $\boldsymbol{W}^{(k)} = \mathcal{W}(\boldsymbol{\Lambda}^{(k-1)}; \boldsymbol{X}_{\mathrm{opt}}, \boldsymbol{R})$
  $\delta_k = t - \|\boldsymbol{X}_{\mathrm{opt}} - \boldsymbol{\Lambda}^{(k)}\boldsymbol{R}\boldsymbol{W}^{(k)}\|_F$
end for
return $\widehat{\boldsymbol{\Lambda}}_{\mathrm{opt}} = \boldsymbol{\Lambda}^{(k)}$, $\widehat{\boldsymbol{W}}_{\mathrm{opt}} = \boldsymbol{W}^{(k)}$

Proposition 2: The sequence $\left\{p_k = \|\boldsymbol{X}_{\mathrm{opt}} - \boldsymbol{\Lambda}^{(k)}\boldsymbol{R}\boldsymbol{W}^{(k)}\|_F\right\}$ is guaranteed to converge to a real number $p^\star \in \mathbb{R}$ that is the objective value achieved by Algorithm 1, and the condition $\delta_k = p_{k-1} - p_k < \epsilon$ is attained in a finite number of iterations.

Proof: The objective value p_k at the kth iteration satisfies

$$\begin{aligned} 0 \leq p_k &= \|\boldsymbol{X}_{\mathrm{opt}} - \boldsymbol{\Lambda}^{(k)}\boldsymbol{R}\boldsymbol{W}^{(k)}\|_F \leq \|\boldsymbol{X}_{\mathrm{opt}} - \boldsymbol{\Lambda}^{(k)}\boldsymbol{R}\boldsymbol{W}^{(k-1)}\|_F \\ &\leq \|\boldsymbol{X}_{\mathrm{opt}} - \boldsymbol{\Lambda}^{(k-1)}\boldsymbol{R}\boldsymbol{W}^{(k-1)}\|_F = p_{k-1}. \end{aligned} \tag{28}$$

Therefore, by the monotone convergence theorem, $\{p_k\}$ is a converging sequence with the limit

$$\lim_{k\to\infty} p_k = p^\star, \tag{29}$$

which is obviously the objective value achieved by Algorithm 1. Moreover, $\{p_k\}$ is a Cauchy sequence since $\mathbb{R}$ has finite dimensions [43], which indicates that for $\forall\epsilon > 0$, $\exists N_\epsilon$, a positive integer, such that $|p_m - p_n| < \epsilon$ if $m, n > N_\epsilon$. Consequently, $\delta_k = p_{k-1} - p_k < \epsilon$ is attained in at most $N_\epsilon + 2$ iterations. ■

IV. SIMULATION STUDY

A mmWave wireless system is simulated, where the transmitter deploys a $\sqrt{N_{\mathrm{t}}} \times \sqrt{N_{\mathrm{t}}}$ UPA with $N_{\mathrm{t}}^{\mathrm{RF}}$ RF chains and the receiver employs a $\sqrt{N_{\mathrm{r}}} \times \sqrt{N_{\mathrm{r}}}$ UPA to support N_{s} independent data streams. The system's signal to noise ratio (SNR) is defined as $\mathrm{SNR} = \frac{\rho}{\sigma^2}$. All the results are obtained by averaging over 10,000 realizations of the channel $\boldsymbol{H}$, which is simulated according to the model (8). The performance of our proposed hybrid precoding architecture based on MRAs (labeled as 'Proposed scheme with MRAs') is compared with those of the three benchmark precoding schemes, the unconstrained optimal full digital precoding solution $\boldsymbol{X}_{\mathrm{opt}}$ (labeled as 'Optimal [13]'), the FC-PAAs scheme with orthogonal matching pursuit (OMP) (labeled as 'FC-PAAs [13]') and the SC-PAAs scheme with near-optimal precoding (labeled as 'SC-PAAs'). The carrier frequency is 28 GHz.

The first two benchmarks are detailed in [13]. We now provide the SC-PAAs solution adopted in this simulation study. Without loss of generality, the channel matrix is uniformly divided into the $N_{\mathrm{t}}^{\mathrm{RF}}$ sub-channels, $\boldsymbol{H} = \left[\boldsymbol{H}_1\, \boldsymbol{H}_2 \cdots \boldsymbol{H}_{N_{\mathrm{t}}^{\mathrm{RF}}}\right]$, and each sub-channel corresponds to one RF chain, which is connected with a subgroup of antenna elements to transmit one data stream. For each sub-channel, the unconstrained optimal precoder is a complex vector $\boldsymbol{f}_i^{\mathrm{opt}} \in \mathbb{C}^{N_{\mathrm{g}}}$, where $N_{\mathrm{g}} = N_{\mathrm{t}}/N_{\mathrm{t}}^{\mathrm{RF}}$. Thus, for the SC-PAAs scheme, the near-optimal analog precoder is given by

$$\boldsymbol{F} = \sqrt{N_{\mathrm{g}}}\begin{bmatrix} \boldsymbol{f}_1^{\mathrm{opt}} \oslash |\boldsymbol{f}_1^{\mathrm{opt}}| & \boldsymbol{0}_{N_{\mathrm{g}}} & \cdots & \boldsymbol{0}_{N_{\mathrm{g}}} \\ \boldsymbol{0}_{N_{\mathrm{g}}} & \boldsymbol{f}_2^{\mathrm{opt}} \oslash |\boldsymbol{f}_2^{\mathrm{opt}}| & \cdots & \boldsymbol{0}_{N_{\mathrm{g}}} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0}_{N_{\mathrm{g}}} & \boldsymbol{0}_{N_{\mathrm{g}}} & \cdots & \boldsymbol{f}_{N_{\mathrm{t}}^{\mathrm{RF}}}^{\mathrm{opt}} \oslash |\boldsymbol{f}_{N_{\mathrm{t}}^{\mathrm{RF}}}^{\mathrm{opt}}| \end{bmatrix}, \tag{30}$$

where each element of $|\boldsymbol{f}_i^{\mathrm{opt}}|$ is the magnitude of the corresponding element in $\boldsymbol{f}_i^{\mathrm{opt}}$, and $\oslash$ denotes the Hadamard division, while the near-optimal digital precoder is given by

$$\boldsymbol{W} = \frac{\sqrt{N_{\mathrm{s}}}}{\left\|\boldsymbol{F}\left(\boldsymbol{F}^{\mathrm{H}}\boldsymbol{F}\right)^{-1}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{X}_{\mathrm{opt}}\right\|_F}\left(\boldsymbol{F}^{\mathrm{H}}\boldsymbol{F}\right)^{-1}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{X}_{\mathrm{opt}}. \tag{31}$$

For the proposed hybrid precoding scheme with MRAs, we set $z/D = 1.2$, $R/D = 0.1$, $\theta = 15°$, $\alpha = 45°$ and $\beta = 30°$. The amplitude variation is determined as $|R_{i,k}| = \cos^q(\Omega_{i,k})$, where $\Omega_{i,k} = \arccos\left(z_k^{\mathrm{RF}}/L_{i,k}\right)$ with $q = 6.5$ [24]. The suboptimal analog and digital precoders, $\widehat{\boldsymbol{\Lambda}}_{\mathrm{opt}}$ and $\widehat{\boldsymbol{W}}_{\mathrm{opt}}$, are obtained by Algorithm 1.

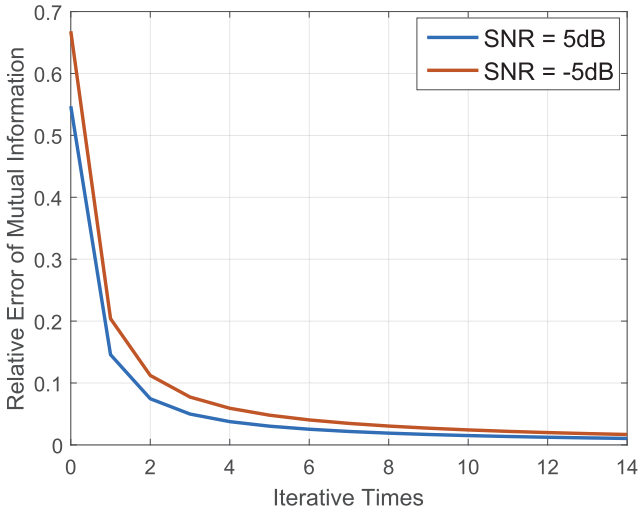

FIGURE 3. Convergence performance of Algorithm 1 given two different values of SNR, where $N_t = 8 \times 8$, $N_r = 4 \times 4$, $N_t^{RF} = 8$, and $N_s = 4$.

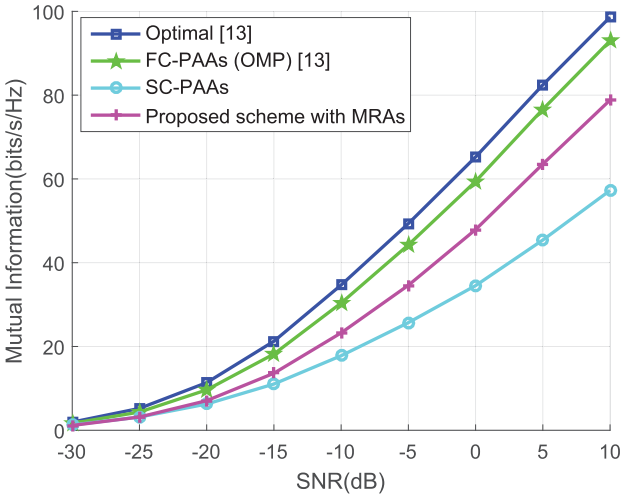

FIGURE 5. Precoding performance of various precoding designs as the functions of the system's SNR, where $N_t = 16 \times 16$, $N_r = 8 \times 8$, $N_t^{RF} = 16$ and $N_s = 10$.

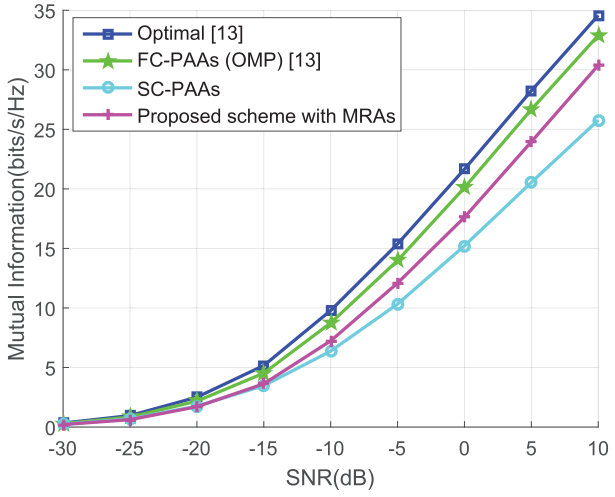

FIGURE 4. Precoding performance of various precoding designs as the functions of the system's SNR, where $N_t = 8 \times 8$, $N_r = 4 \times 4$, $N_t^{RF} = 8$ and $N_s = 4$.

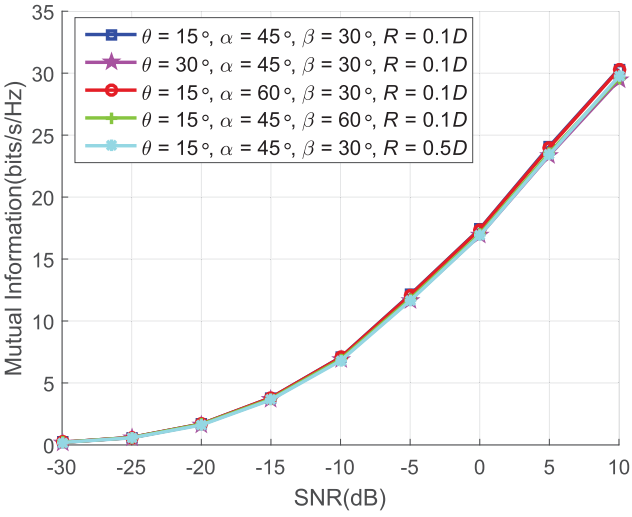

FIGURE 6. Precoding performance is insensitive to the transfer matrix R as long as the RF feeds are configured within a reasonable spatial scope. The transceiver parameters are $N_t = 8 \times 8$, $N_r = 4 \times 4$, $N_t^{RF} = 8$ and $N_s = 4$.

A. PRECODING PERFORMANCE

First, the convergence performance of Algorithm 1 for two different values of SNR are demonstrated in Fig. 3, where the relative error of the MI in the kth iteration is defined as $\left|\frac{I\left(\mathbf{\Lambda}^{(k)}, \boldsymbol{W}^{(k)}\right) - p^\star}{p^\star}\right|$, with $p^\star$ set to $I\left(\mathbf{\Lambda}^{(200)}, \boldsymbol{W}^{(200)}\right)$, which is a sufficiently accurate approximation of the optimal MI value that the algorithm converges to. Observe that it takes 5 iterations on average to reduce the relative error to below 5%. As expected, better convergence performance is achieved with higher SNR.

The performance of the four precoding designs as the functions of the system's SNR are compared in Figs. 4 and 5 for the two sets of the transceiver's parameters, respectively. Observe that the performance of the FC-PAAs based design is very close to that of the optimal full precoding solution $\boldsymbol{X}_{\text{opt}}$ but it suffers from the serious drawback of imposing very high hardware complexity, as shown in Table 1. Compared with the FC-PAAs based design, the SC-PAAs based design offers considerably lower hardware complexity, as can be seen from Table 1, but suffers from serious performance loss, as shown in Figs. 4 and 5. Remarkably, our proposed hybrid precoding design based on MRAs, which has an even lower hardware complexity than the SC-PAAs based design, outperforms the SC-PAAs based design considerably, although it suffers from some performance loss in comparison with the FC-PAAs based design whose very high hardware requirements may prevent its practical implementation. Our results therefore demonstrate that the MRAs based precoding with its remarkable hardware efficiency is capable of providing a practical solution to mmWave wireless systems with acceptably good performance.

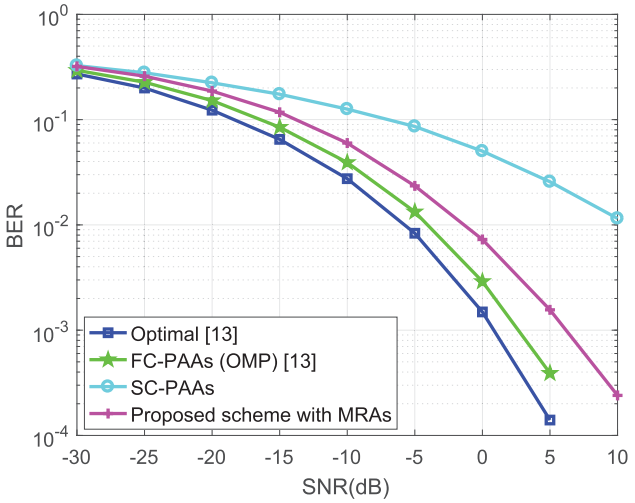

FIGURE 7. System BER performance achieved by various precoding designs as the functions of the system's SNR, where $N_t = 8 \times 8$, $N_r = 4 \times 4$, $N_t^{RF} = 8$ and $N_s = 4$, while the full-digital ZF combiner is employed at the receiver.

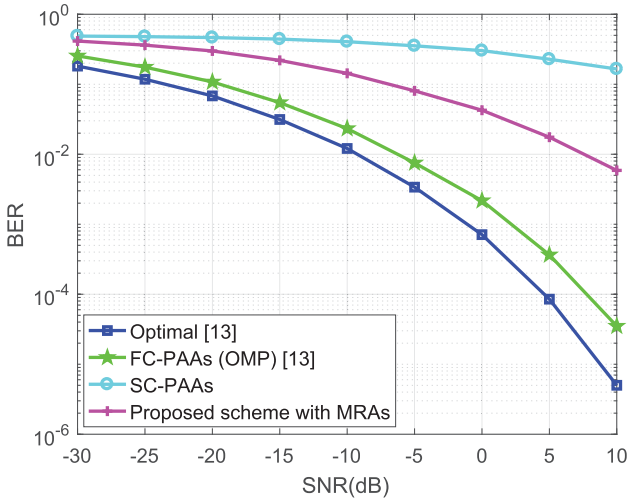

FIGURE 8. System BER performance achieved by various precoding designs as the functions of the system's SNR, where $N_t = 16 \times 16$, $N_r = 8 \times 8$, $N_t^{RF} = 16$ and $N_s = 10$, while the full-digital ZF combiner is employed at the receiver.

Next, the results of Fig. 6 confirm that the precoding performance of our MRAs based precoding design is insensitive to the transfer matrix $\boldsymbol{R}$ as long as the RF feeds are configured within a reasonable spatial scope.

B. SYSTEM PERFORMANCE

To compare the bit error rate (BER) performance achieved by various precoding designs, we utilize the full digital combiner at the receiver, with the detected symbol vector given by

$$\widehat{\boldsymbol{s}} = \boldsymbol{P}\boldsymbol{y} = \boldsymbol{P}\boldsymbol{H}_{\mathrm{E}}\boldsymbol{s} + \boldsymbol{P}\boldsymbol{\xi}, \tag{32}$$

where $\boldsymbol{P} \in \mathbb{C}^{N_s \times N_r}$ is the full combining matrix, and $\boldsymbol{H}_{\mathrm{E}} = \sqrt{\rho}\boldsymbol{H}\boldsymbol{X} \in \mathbb{C}^{N_r \times N_s}$ is the equivalent channel matrix in which $\boldsymbol{X}$ is the full precoding matrix deployed at the transmitter. In the simulation, we use the full digital zero-forcing (ZF) combining of the form

$$\boldsymbol{P}_{\mathrm{ZF}} = \left(\boldsymbol{H}_{\mathrm{E}}^{\mathrm{H}}\boldsymbol{H}_{\mathrm{E}}\right)^{-1}\boldsymbol{H}_{\mathrm{E}}^{\mathrm{H}}. \tag{33}$$

By adopting the same full digital combiner at the receiver, we can compare the system's achievable BER performance for different precoding designs employed at the transmitter.

The achievable system's BER performance for the four precoding designs as the functions of the system's SNR are compared in Figs. 7 and 8 given the two sets of the transceiver's parameters, respectively. As expected, in line with the achievable precoding performance shown in Figs. 4 and 5, it can be observed that our proposed hybrid precoding design based on MRAs attains much better BER performance than the SC-PAAs based design, and yet it imposes a lower hardware complexity than the latter.

V. CONCLUSIONS

A hybrid precoding design based on multi-feed reflectarray antennas has been proposed for large-scale antenna array aided mmWave wireless systems to reduce the hardware implementation complexity significantly while attaining acceptably good performance. We have formulated the mathematical model for jointly designing the analog and digital precoders, and have derived an efficient iterative algorithm to obtain suboptimal solution, which guarantees to converge fast. Our results have demonstrated that the proposed hybrid precoding scheme based on MRAs outperforms the widely adopted low-hardware-complexity SC-PAAs based scheme both in terms of precoding performance and hardware requirements, and it offers a practical and attractive tradeoff between achievable precoding performance and hardware complexity over the FC-PAAs based design of very high hardware requirements.

APPENDIX

A. PROOF OF LEMMA 1

Proof: We can express $\mathrm{tr}\left(\boldsymbol{D}^{\mathrm{H}}\boldsymbol{D}\right)$ as

$$\begin{aligned}\mathrm{tr}\left(\boldsymbol{D}^{\mathrm{H}}\boldsymbol{D}\right) &= \mathrm{tr}\left(\left(\boldsymbol{A}\boldsymbol{P}^{\mathrm{H}} + \boldsymbol{P}\boldsymbol{A}^{\mathrm{H}}\right)\left(\boldsymbol{A}\boldsymbol{P}^{\mathrm{H}} + \boldsymbol{P}\boldsymbol{A}^{\mathrm{H}}\right)\right) \\ &= \mathrm{tr}\left(\boldsymbol{A}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{P}^{\mathrm{H}}\right) + \mathrm{tr}\left(\boldsymbol{P}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{P}\boldsymbol{A}^{\mathrm{H}}\right) \\ &\quad + \mathrm{tr}\left(\boldsymbol{A}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P}\boldsymbol{A}^{\mathrm{H}}\right) + \mathrm{tr}\left(\boldsymbol{P}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{P}^{\mathrm{H}}\right).\end{aligned} \tag{34}$$

As proved in [44, Th. 1], both $\mathrm{tr}\left(\boldsymbol{A}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P}\boldsymbol{A}^{\mathrm{H}}\right)$ and $\mathrm{tr}\left(\boldsymbol{P}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{P}^{\mathrm{H}}\right)$ are bounded, specifically,

$$\begin{aligned}\mathrm{tr}\left(\boldsymbol{A}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P}\boldsymbol{A}^{\mathrm{H}}\right) &\le \mathrm{tr}\left(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\right)\mathrm{tr}\left(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P}\right), \\ \mathrm{tr}\left(\boldsymbol{P}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{P}^{\mathrm{H}}\right) &\le \mathrm{tr}\left(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\right)\mathrm{tr}\left(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P}\right).\end{aligned} \tag{35}$$

As for $\mathrm{tr}\left(\boldsymbol{A}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{P}^{\mathrm{H}}\right)$, by denoting $\boldsymbol{S} = \boldsymbol{A}\boldsymbol{P}^{\mathrm{H}}$, we have

$$\begin{aligned}\left|\mathrm{tr}\left(\boldsymbol{A}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{P}^{\mathrm{H}}\right)\right| &= \left|\sum_{i=1}^{N}\sum_{k=1}^{N}\boldsymbol{S}_{i,k}\boldsymbol{S}_{k,i}\right| \\ &\le \sqrt{\sum_{i=1}^{N}\sum_{k=1}^{N}\left|\boldsymbol{S}_{i,k}\right|^2\sum_{i=1}^{N}\sum_{k=1}^{N}\left|\boldsymbol{S}_{k,i}\right|^2} \\ &= \mathrm{tr}\left(\boldsymbol{P}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{P}^{\mathrm{H}}\right) \\ &\le \mathrm{tr}\left(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\right)\mathrm{tr}\left(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P}\right).\end{aligned} \tag{36}$$

Similarly, we have

$$\left|\mathrm{tr}\left(\boldsymbol{P}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{P}\boldsymbol{A}^{\mathrm{H}}\right)\right| \leq \mathrm{tr}\left(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\right)\mathrm{tr}\left(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P}\right). \tag{37}$$

This completes the proof. ■

B. PROOF OF LEMMA 2

Proof: Based on the well-known properties of eigenvalues and singular values, we have

$$\begin{aligned}\left|\lambda_i\left(\boldsymbol{H}\boldsymbol{D}\boldsymbol{H}^{\mathrm{H}}\right)\right| &\leq \sigma_1\left(\boldsymbol{H}\boldsymbol{D}\boldsymbol{H}^{\mathrm{H}}\right)\\ &\leq \frac{1}{2}\left(1+\lambda_1\left(\boldsymbol{H}\boldsymbol{D}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{D}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\right)\right)\\ &\leq \frac{1}{2}\left(1+\mathrm{tr}\left(\boldsymbol{H}\boldsymbol{D}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{D}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\right)\right)\\ &\leq \frac{1}{2}+\frac{1}{2}\mathrm{tr}\left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\right)\mathrm{tr}\left(\boldsymbol{D}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{D}^{\mathrm{H}}\right)\\ &\leq \frac{1}{2}+\frac{1}{2}\left(\mathrm{tr}\left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\right)\right)^2\mathrm{tr}\left(\boldsymbol{D}^{\mathrm{H}}\boldsymbol{D}\right)\\ &\triangleq \nu,\end{aligned} \tag{38}$$

where the last two inequalities hold according to [44, Th. 1]. This completes the proof. ■

C. PROOF OF PROPOSITION 1

Proof: Given a fixed point $\boldsymbol{P} \in \mathbb{C}^{N_{\mathrm{t}}\times N_{\mathrm{s}}}$ in the metric space $\mathcal{B}\left(\mathbb{C}^{N_{\mathrm{t}}\times N_{\mathrm{s}}}, \|\cdot\|_F\right)$ and $\boldsymbol{H} \in \mathbb{C}^{N_{\mathrm{r}}\times N_{\mathrm{t}}}$, for any $\boldsymbol{X} \in \mathbb{C}^{N_{\mathrm{t}}\times N_{\mathrm{s}}}$ such that $\|\boldsymbol{X}-\boldsymbol{P}\|_F^2 = \delta > 0$, without loss of generally, $\boldsymbol{X}$ can be expressed as,

$$\boldsymbol{X} = \boldsymbol{P} + \sqrt{\delta}\boldsymbol{A}, \tag{39}$$

where $\boldsymbol{A} \in \mathbb{C}^{N_{\mathrm{t}}\times N_{\mathrm{s}}}$ is any normalized matrix such that $\mathrm{tr}\left(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\right) = 1$. Then

$$\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}} = \boldsymbol{P}\boldsymbol{P}^{\mathrm{H}} + \sqrt{\delta}\boldsymbol{A}\boldsymbol{P}^{\mathrm{H}} + \sqrt{\delta}\boldsymbol{P}\boldsymbol{A}^{\mathrm{H}} + \delta\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}. \tag{40}$$

If we denote $\boldsymbol{T} = \boldsymbol{H}\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}$, $\boldsymbol{V} = \boldsymbol{H}\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}$, $\boldsymbol{U} = \boldsymbol{H}\left(\boldsymbol{A}\boldsymbol{P}^{\mathrm{H}} + \boldsymbol{P}\boldsymbol{A}^{\mathrm{H}}\right)\boldsymbol{H}^{\mathrm{H}}$, and $\boldsymbol{Z} = \boldsymbol{H}\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}$, according to Weyl's theorem [45], we have the following two inequalities

$$\lambda_i\left(\boldsymbol{T}\right) \geq \lambda_i\left(\boldsymbol{V}\right) + \sqrt{\delta}\lambda_{N_{\mathrm{r}}}\left(\boldsymbol{U}\right) + \delta\lambda_{N_{\mathrm{r}}}\left(\boldsymbol{Z}\right), \tag{41}$$

$$\lambda_i\left(\boldsymbol{T}\right) \leq \lambda_i\left(\boldsymbol{V}\right) + \sqrt{\delta}\lambda_1\left(\boldsymbol{U}\right) + \delta\lambda_1\left(\boldsymbol{Z}\right). \tag{42}$$

According to Lemmas 1 and 2, $\lambda_i\left(\boldsymbol{U}\right)$ is bounded by $\nu_{\boldsymbol{U}}$ and $\lambda_i\left(\boldsymbol{Z}\right)$ is bounded by $\nu_{\boldsymbol{Z}}$, with

$$\nu_{\boldsymbol{U}} = \frac{1}{2} + 2\mathrm{tr}\left(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P}\right)\left(\mathrm{tr}\left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\right)\right)^2, \tag{43}$$

$$\nu_{\boldsymbol{Z}} = \frac{1}{2} + \frac{1}{2}\left(\mathrm{tr}\left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\right)\right)^2. \tag{44}$$

Hence we have

$$\lambda_L(\delta) \leq \lambda_i\left(\boldsymbol{T}\right) \leq \lambda_U(\delta), \tag{45}$$

where

$$\lambda_L(\delta) = \lambda_i\left(\boldsymbol{V}\right) - \sqrt{\delta}\nu_{\boldsymbol{U}} - \delta\nu_{\boldsymbol{Z}}, \tag{46}$$

$$\lambda_U(\delta) = \lambda_i\left(\boldsymbol{V}\right) + \sqrt{\delta}\nu_{\boldsymbol{U}} + \delta\nu_{\boldsymbol{Z}}. \tag{47}$$

Notice that $\lambda_L(\delta)$ is positive for sufficiently small δ. On the other hand,

$$\begin{aligned}\mathrm{tr}\left((\boldsymbol{A}-\boldsymbol{P})^{\mathrm{H}}(\boldsymbol{A}-\boldsymbol{P})\right) &= \mathrm{tr}\left(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\right) + \mathrm{tr}\left(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P}\right)\\ &\quad - \mathrm{tr}\left(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{P}\right) - \mathrm{tr}\left(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{A}\right) \geq 0.\end{aligned} \tag{48}$$

Thus, we have

$$\mathrm{tr}\left(\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\right) \leq \left(1+\sqrt{\delta}\right)\left(\mathrm{tr}\left(\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}\right) + \sqrt{\delta}\mathrm{tr}\left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\right)\right). \tag{49}$$

As $\delta \to 0$, the lower bound of $\bar{I}(\boldsymbol{X})$ is given by

$$\begin{aligned}\bar{I}(\boldsymbol{X}) &= \log_2\left(\det\left(\boldsymbol{I}_{N_{\mathrm{r}}} + \frac{1}{\mathrm{tr}\left(\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\right)}\boldsymbol{H}\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\right)\right)\\ &= \log_2\left(\prod_{i=1}^{N_{\mathrm{r}}}\left|1+\frac{\lambda_i\left(\boldsymbol{H}\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\right)}{\mathrm{tr}\left(\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\right)}\right|\right)\\ &\geq \log_2\left(\prod_{i=1}^{N_{\mathrm{r}}}\left|1+\frac{\lambda_i\left(\boldsymbol{H}\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\right)}{\left(1+\sqrt{\delta}\right)\left(\mathrm{tr}\left(\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}\right)+\sqrt{\delta}\mathrm{tr}\left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\right)\right)}\right|\right)\\ &\geq \log_2\left(\prod_{i=1}^{N_{\mathrm{r}}}\left|1+\frac{\lambda_i\left(\boldsymbol{H}\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\right)-\sqrt{\delta}\nu_{\boldsymbol{U}}-\delta\nu_{\boldsymbol{Z}}}{\left(1+\sqrt{\delta}\right)\left(\mathrm{tr}\left(\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}\right)+\sqrt{\delta}\right)}\right|\right)\\ &\to \bar{I}(\boldsymbol{P}),\end{aligned} \tag{50}$$

while the upper bound of $\bar{I}(\boldsymbol{X})$ is expressed as

$$\begin{aligned}\bar{I}(\boldsymbol{X}) &= \log_2\left(\prod_{i=1}^{N_{\mathrm{r}}}\left|1+\frac{\lambda_i\left(\boldsymbol{H}\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\right)}{\mathrm{tr}\left(\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\right)}\right|\right)\\ &\leq \log_2\left(\prod_{i=1}^{N_{\mathrm{r}}}\left|1+\frac{\lambda_i\left(\boldsymbol{H}\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\right)+\sqrt{\delta}\nu_{\boldsymbol{U}}+\delta\nu_{\boldsymbol{Z}}}{\mathrm{tr}\left(\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\right)}\right|\right)\\ &= \log_2\left(\prod_{i=1}^{N_{\mathrm{r}}}\left|1+\frac{\lambda_i\left(\boldsymbol{H}\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\right)+\sqrt{\delta}\nu_{\boldsymbol{U}}+\delta\nu_{\boldsymbol{Z}}}{\mathrm{tr}\left(\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}\right)+o(\delta)}\right|\right)\\ &\to \bar{I}(\boldsymbol{P}) = \log_2\left(\prod_{i=1}^{N_{\mathrm{r}}}\left|1+\frac{\lambda_i\left(\boldsymbol{H}\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\right)}{\mathrm{tr}\left(\boldsymbol{P}\boldsymbol{P}^{\mathrm{H}}\right)}\right|\right),\end{aligned} \tag{51}$$

where

$$o(\delta) = \sqrt{\delta}\mathrm{tr}\left(\boldsymbol{A}\boldsymbol{P}^{\mathrm{H}} + \boldsymbol{P}\boldsymbol{A}^{\mathrm{H}}\right) + \delta \to 0 \text{ as } \delta \to 0, \tag{52}$$

because $\mathrm{tr}\left(\boldsymbol{A}\boldsymbol{P}^{\mathrm{H}} + \boldsymbol{P}\boldsymbol{A}^{\mathrm{H}}\right)$ in $o(\delta)$ is bounded:

$$\begin{aligned}\left|\mathrm{tr}\left(\boldsymbol{A}\boldsymbol{P}^{\mathrm{H}}\right)\right| &= \left|\sum_{i=1}^{N_{\mathrm{r}}}\sum_{k=1}^{N_{\mathrm{r}}}\boldsymbol{A}_{i,k}\boldsymbol{P}_{k,i}^{\mathrm{H}}\right|\\ &\leq \sqrt{\sum_{i=1}^{N_{\mathrm{r}}}\sum_{k=1}^{N_{\mathrm{r}}}\left|\boldsymbol{A}_{i,k}\right|^2\sum_{i=1}^{N_{\mathrm{r}}}\sum_{k=1}^{N_{\mathrm{r}}}\left|\boldsymbol{P}_{k,i}^{\mathrm{H}}\right|^2}\\ &= \sqrt{\mathrm{tr}\left(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\right)\mathrm{tr}\left(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P}\right)}.\end{aligned} \tag{53}$$

Therefore, $\bar{I}(\boldsymbol{X})$ is continuous at $\forall \boldsymbol{P} \in \mathcal{B}\left(\mathbb{C}^{N_{\mathrm{t}}\times N_{\mathrm{s}}}, \|\cdot\|_F\right)$. This completes the proof. ■

REFERENCES

[1] A. Ghosh *et al.*, "Millimeter-wave enhanced local area systems: A high-data-rate approach for future wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1152–1163, Jun. 2014.

[2] 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.

[3] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1617–1655, 3rd Quart., 2016.

[4] 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, Mar. 2014.

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

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

[7] Y. Jiang, J. Li, and W. M. Hager, "Uniform channel decomposition for MIMO communications," *IEEE Trans. Signals Process.*, vol. 53, no. 11, pp. 4283–4294, Nov. 2005.

[8] W. Yao, S. Chen, and L. Hanzo, "A transceiver design based on uniform channel decomposition and MBER vector perturbation," *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 3153–3159, Jul. 2010.

[9] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[10] 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.

[11] V. Venkateswaran and A. van der Veen, "Analog beamforming in MIMO communications with phase shift networks and online channel estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4131–4143, Aug. 2010.

[12] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "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.

[13] 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.

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

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

[16] 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.

[17] V. Venkateswaran, F. Pivit, and L. Guan, "Hybrid RF and digital beamformer for cellular networks: Algorithms, microwave architectures, and measurements," *IEEE Transa. Microw. Theory Techn.*, vol. 64, no. 7, pp. 2226–2243, Jul. 2016.

[18] A. Garcia-Rodriguez, V. Venkateswaran, P. Rulikowski, and C. Masouros, "Hybrid analog–digital precoding revisited under realistic RF modeling," *IEEE Wireless Commun. Lett.*, vol. 5, no. 5, pp. 528–531, Oct. 2016.

[19] J. Huang and J. A. Encinar, *Reflectarray Antennas*. Hoboken, NJ, USA: Wiley, 2008.

[20] P. Nayeri, F. Yang, and A. Z. Elsherbeni, "Single-feed multi-beam reflectarray antennas," in *Proc. APSURSI*, Toronto, ON, Canada, Jul. 2010, pp. 1–4.

[21] F. Yang, "Progress in reflectarray antenna research: From enhanced frequency features to advanced radiation capabilities," in *Proc. EUCAP*, Gothenburg, Sweden, Apr. 2013, pp. 2484–2487.

[22] K. Lele, A. A. Desai, A. A. Kadam, and A. A. Deshmukh, "Reflectarray antennas," *Int. J. Comput. Appl.*, vol. 108, no. 3, pp. 21–28, Dec. 2014.

[23] Y. Qu, C. Guo, H. Guo, and J. Ding, "Design of single-feed multi-beam reflectarray using iterative Fourier techniques," *Int. J. Smart Home*, vol. 9, no. 4, pp. 187–194, 2015.

[24] P. Nayeri, F. Yang, and A. Z. Elsherbeni, "Design of single-feed reflectarray antennas with asymmetric multiple beams using the particle swarm optimization method," *IEEE Trans. Antennas Propag.*, vol. 61, no. 9, pp. 4598–4605, Sep. 2013.

[25] M. Arrebola, J. A. Encinar, and M. Barba, "Multifed printed reflectarray with three simultaneous shaped beams for LMDS central station antenna," *IEEE Trans. Antennas Propag.*, vol. 56, no. 6, pp. 1518–1527, Jun. 2008.

[26] P. Nayeri, F. Yang, and A. Z. Elsherbeni, "Beam-scanning reflectarray antennas: A technical overview and state of the art," *IEEE Antennas Propag. Mag.*, vol. 57, no. 4, pp. 32–47, Aug. 2015.

[27] J. Ethier, M. R. Chaharmir, J. Shaker, and D. Lee, "Development of novel low-cost reflectarrays [antenna applications corner]," *IEEE Antennas Propag. Mag.*, vol. 54, no. 3, pp. 277–287, Jun. 2012.

[28] W. Lee and Y. J. Yoon, "A broadband dual-metallic-reflectarray antenna for millimeter-wave applications," *IEEE Antennas Wireless Propag. Lett.*, vol. 16, pp. 856–859, 2017.

[29] M. H. Dahri, M. H. Jamaluddin, M. I. Abbasi, and M. R. Kamarudin, "A review of wideband reflectarray antennas for 5G communication systems," *IEEE Access*, vol. 5, pp. 17803–17815, 2017.

[30] A. Tamminen *et al.*, "Reflectarray design for 120-GHz radar application: Measurement results," *IEEE Trans. Antennas Propag.*, vol. 61, no. 10, pp. 5036–5047, Oct. 2013.

[31] Q. Luo *et al.*, "Design and analysis of a reflectarray using slot antenna elements for Ka-band SatCom," *IEEE Trans. Antennas Propag.*, vol. 62, no. 4, pp. 1365–1374, Apr. 2015.

[32] O. M. Haraz and M. M. M. Ali, "A millimeter-wave circular reflectarray antenna for future 5G cellular networks," in *Proc. IEEE Int. Symp. Antennas Propag. USNC-URSI Nat. Radio Sci. Meeting*, Vancouver, BC, Canada, Jul. 2015, pp. 1534–1535.

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

[34] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6169–6183, Dec. 2015.

[35] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels," *IEEE Commun. Let.*, vol. 20, no. 6, pp. 1259–1262, Jun. 2016.

[36] Z. Wang, P. Zhao, C. Qian, and S. Chen, "Location-aware channel estimation enhanced TDD based massive MIMO," *IEEE Access*, vol. 4, pp. 7828–7840, Nov. 2016.

[37] X. Guo, S. Chen, J. Zhang, X. Mu, and L. Hanzo, "Optimal pilot design for pilot contamination elimination/reduction in large-scale multiple-antenna aided OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7229–7243, Nov. 2016.

[38] S.-J. Kim, A. Magnani, A. Mutapcic, S. P. Boyd, and Z.-Q. Luo, "Robust beamforming via worst-case SINR maximization," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1539–1547, Apr. 2008.

[39] M. Ding and S. D. Blostein, "MIMO minimum total MSE transceiver design with imperfect CSI at both ends," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1141–1150, Mar. 2009.

[40] C. Xing, Z. Fei, S. Ma, Y. C. Wu, and H. V. Poor, "A general robust linear transceiver design for multi-hop amplify-and-forward MIMO relaying systems," *IEEE Trans. Signal Process.*, vol. 61, no. 5, pp. 1196–1209, Mar. 2013.

[41] S. Gong, C. Xing, S. Chen, and Z. Fei, "Secure communications for dual-polarized MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 16, pp. 4177–4192, Aug. 2017.

[42] S. Gong, C. Xing, S. Chen, N. Yang, and Y. Zhou, "Robust energy-efficient precoding optimization for dual-polarized multiuser MIMO downlink," presented at the ICCC, Chengdu, China, Dec. 2017, pp. 1–5.

[43] L. H. Royden, *Real Analysis*. New York, NY, USA: Macmillan, 1988.

[44] I. D. Coope, "On matrix trace inequalities and related topics for products of Hermitian matrices," *J. Math. Anal. Appl.*, vol. 188, no. 3, pp. 999–1001, Dec. 1994.
[45] P. Lancaster and M. Tismenetsky, *The Theory of Matrices: With Applications*. New York, NY, USA: Academic, 1985.

ZHENGYI ZHOU received the B.S. degree in electronic engineering from Tsinghua University, China, in 2016, where he is currently pursuing the Ph.D. degree in communications and information systems. His research interests include MIMO, precoding, and antenna systems.

NING GE (M'97) received the B.S. and Ph.D. degrees from Tsinghua University, China, in 1993 and 1997, respectively. From 1998 to 2000, he was involved in the development of ATM switch fabric ASIC with ADC Telecommunications, Dallas. Since 2000, he has been with the Department of Electronics Engineering, Tsinghua University, where he is currently a Full Professor and also serves as the Director of Communication Institute. His research interests include ASIC design, short-range wireless communication, and wireless communications. He is a Senior Member of CIC and CIE.

ZHAOCHENG WANG (M'09–SM'11) received the B.S., M.S., and Ph.D. degrees from Tsinghua University in 1991, 1993, and 1996, respectively. From 1996 to 1997, he was a Post-Doctoral Fellow with Nanyang Technological University, Singapore. From 1997 to 1999, he was a Research Engineer/Senior Engineer with OKI Techno Centre (Singapore) Pte. Ltd., Singapore. From 1999 to 2009, he was a Senior Engineer/Principal Engineer with Sony Deutschland GmbH, Germany. Since 2009, he has been a Full Professor with the Department of Electronic Engineering, Tsinghua University, where he currently serves as the Director of Broadband Communication Key Laboratory and the Tsinghua National Laboratory for Information Science and Technology. He has authored around 200 peer-reviewed journal and conference papers. He has co-authored two books, which were selected by the IEEE Series on Digital and Mobile Communication and published by Wiley–IEEE Press. He holds 34 U.S./EU granted patents (23 of them as the first inventor and 9 of them as the second inventor). His research interests include wireless communications, optical wireless communications, and millimeter wave communications. He is a Fellow of the Institution of Engineering and Technology. He has received the ICC2013 Best Paper Award, the OECC2015 Best Student Paper Award, the 2016 IEEE Scott Helt Memorial Award (Best Paper Award of IEEE TRANSACTIONS ON BROADCASTING), and the ICC2017 Best Paper Award. He has also served as the Technical Program Co-Chair of many international conferences, including ICC and GlobeSIP. He served as an Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2011 to 2015 and the IEEE COMMUNICATIONS LETTERS from 2013 to 2016.

SHENG CHEN (M'90–SM'97–F'08) received the B. Eng. degree in control engineering from the East China Petroleum Institute, Dongying, China, in 1982, the Ph.D. degree in control engineering from City University, London, in 1986, and the D.Sc. degree from the University of Southampton, Southampton, U.K., in 2005. From 1986 to 1999, he held research and academic appointments with The University of Sheffield, U.K., the University of Edinburgh, U.K., and the University of Portsmouth, U.K. Since 1999, he has been with the School of Electronics and Computer Science, University of Southampton, where he is currently a Professor in intelligent systems and signal processing. He is also a Distinguished Adjunct Professor with King Abdulaziz University, Jeddah, Saudi Arabia. He has authored over 550 research papers. His research interests include adaptive signal processing, wireless communications, the modeling and identification of nonlinear systems, neural network and machine learning, intelligent control system design, and evolutionary computation methods and optimisation. He is a fellow of the Royal Academy of Engineering, U.K., a fellow of IET, and an ISI Highly Cited Researcher in engineering in 2004.