Geometric Mean Decomposition Based Hybrid Precoding for Millimeter-Wave Massive MIMO

Tian Xie[1], Linglong Dai[1,*], Xinyu Gao[1], Muhammad Zeeshan Shakir[2], Jianjun Li[3]

[1] Tsinghua National Laboratory for Information Science and Technology (TNList), Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

[2] School of Engineering and Computing, University of the West Scotland (UWS), Glasgow, Scotland

[3] The School of Electric and Information Engineering, Zhongyuan University of Technology, Zhengzhou 450007, China

* The corresponding author, email:daill@tsinghua.edu.cn

Abstract: Hybrid precoding can reduce the number of required radio frequency (RF) chains in millimeter-Wave (mmWave) massive MIMO systems. However, existing hybrid precoding based on singular value decomposition (SVD) requires the complicated bit allocation to match the different signal-to-noise-ratios (SNRs) of different sub-channels. In this paper, we propose a geometric mean decomposition (GMD)-based hybrid precoding to avoid the complicated bit allocation. Specifically, we seek a pair of analog and digital precoders sufficiently close to the unconstrained fully digital GMD precoder. To achieve this, we fix the analog precoder to design the digital precoder, and vice versa. The analog precoder is designed based on the orthogonal matching pursuit (OMP) algorithm, while GMD is used to obtain the digital precoder. Simulations show that the proposed GMD-based hybrid precoding achieves better performance than the conventional SVD-based hybrid precoding with only a slight increase in complexity.

Keywords: Millimeter-wave Massive MIMO; hybrid precoding; geometric mean decomposition; bit allocation

I. Introduction

Millimeter-wave (mmWave) massive MIMO is considered as a promising technology for the future 5G communications [1-6], since it can substantially increase the system throughput. For the sake of low energy consumption and hardware cost, mmWave massive MIMO usually utilizes the hybrid precoding architecture [7-9], where the conventional high-dimensional fully digital precoder is split into a high-dimensional analog precoder and a low-dimensional digital precoder to reduce the number of required radio frequency (RF) chains.

How to obtain the optimal analog and digital precoders is the most important issue for hybrid precoding. In the pioneering work [8], the hybrid precoder design problem was reformulated as a sparse signal reconstruction problem, which is solved via a spatially sparse hybrid precoding algorithm. In [10], an alternating hybrid beamforming approach is designed to optimize the achievable rate of system. In [11], the energy-efficiency oriented hybrid precoding is investigated, where an iterative algorithm is designed to maximize the energy-efficiency of the system. All of these works aim to seek a pair of analog and digital precoders that are sufficiently close to the right singular matrix obtained through the singular value decomposition (SVD) of the channel matrix. Such SVD-based hybrid precoding

Received: Jul. 20, 2017
Revised: Sep. 4, 2017
Editor: Yongming Huang

In this paper, a GMD-based hybrid precoding is proposed to avoid the complicated bit allocation in the conventional SVD-based hybrid precoding.

can achieve the capacity-approaching performance with the optimal water-filling power allocation in point to point-MIMO-systems [8]. However, different sub-channels in SVD-based hybrid precoding usually have different signal-to-noise ratios (SNRs). Therefore, complicated bit allocation, i.e., allocating different modulation and coding schemes (MCSs) on different sub-channels, is usually required. Such procedure involves high coding/decoding complexity in practical systems [12].

In this paper, a geometric mean decomposition (GMD)-based hybrid precoding is proposed to avoid the complicated bit allocation required in the SVD-based hybrid precoding. Unlike the conventional SVD-based hybrid precoding, the right semi-unitary matrix obtained by GMD is treated as the optimal unconstrained precoder, which can convert the mmWave massive MIMO channel into sub-channels with *identical* SNRs [19]. Therefore, the proposed GMD-based hybrid precoding is able to naturally avoid the complicated bit allocation. Then, to find a near-optimal solution to the GMD-based hybrid precoder design problem, which is challenging due to the non-convex constraint on the analog precoder, we use a decoupled optimization method to design the analog and digital precoder. Specifically, the analog precoder design problem is solved via the orthogonal matching pursuit (OMP) algorithm, while the digital precoder is obtained by using GMD. Simulation results show that the proposed GMD-based hybrid precoding achieves better performance than the conventional SVD-based hybrid precoding with only a slight increase in complexity.

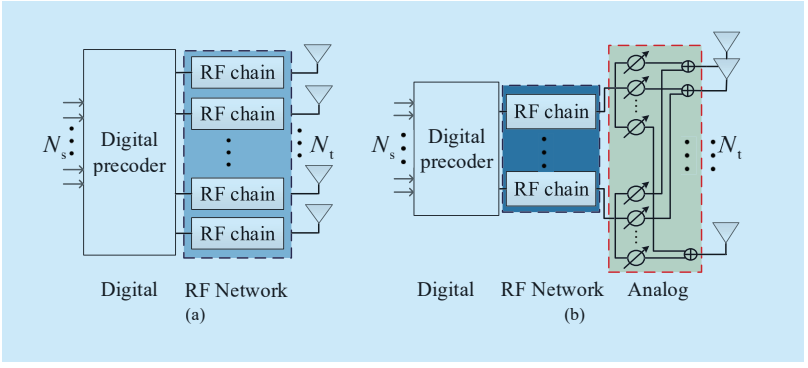


Fig. 1. *Comparison of two typical precoding structures. (a) The fully digital precoding; (b) The hybrid precoding.*

The remainder of this paper is organized as follow. The system model is briefly introduced in Section II. In Section III, we propose the GMD-based hybrid precoding. Then, we evaluate the performance of proposed GMD-based hybrid precoding through simulations in Section IV. Finally, we conclude this paper in Section V.

Notation: $\mathbf{a}$, $\mathbf{A}$, and $\mathcal{A}$ denote a vector, a matrix, and a set. $\mathbb{E}\{\mathbf{A}\}$, $\mathrm{tr}\{\mathbf{A}\}$, $\{\mathbf{A}\}^T$, $\{\mathbf{A}\}^H$, $\{\mathbf{A}\}^\dagger$, $\|\mathbf{A}\|_F$, $\mathbf{A}^{(k)}$, and $\{\mathbf{A}\}_{i,j}$ denote the expectation, trace, transpose, conjugate transpose, pseudo-inverse, Frobenius norm, kth column, and the element in the ith row and the jth column of $\mathbf{A}$, respectively. $\mathbb{C}$ is the set of complex numbers. $\mathbf{I}_{N_s}$ is the $N_s \times N_s$ identity matrix. $|\mathcal{A}|$ is the cardinality of $\mathcal{A}$. Finally, $\mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_s})$ denotes the complex Gaussian distribution with expectation $\mathbf{0}$ and covariance $\mathbf{I}_{N_s}$.

II. System Model

In this section, we introduce the concept of hybrid precoding for mmWave massive MIMO and the widely used mmWave channel model as well.

2.1 Hybrid precoding for mmWave massive MIMO

We consider a typical mmWave massive MIMO system with hybrid precoding, where the base station (BS) with N_t transmit antennas sends N_s independent data streams to the user with N_r receive antennas. We assume that the BS and the user have N_t^{RF} and N_r^{RF} RF chains, respectively, which satisfy $N_s \le N_t^{RF} \le N_t$ and $N_s \le N_r^{RF} \le N_r$ [8]. The received signal $\mathbf{y} \in \mathbb{C}^{N_s \times 1}$ at the user can be expressed as

$$\begin{aligned}\mathbf{y} &= \sqrt{\rho}\mathbf{W}^H\mathbf{HPs} + \mathbf{W}^H\mathbf{n} \\ &= \sqrt{\rho}\mathbf{W}_D^H\mathbf{W}_A^H\mathbf{HP}_A\mathbf{P}_D\mathbf{s} + \mathbf{W}_D^H\mathbf{W}_A^H\mathbf{n},\end{aligned} \quad (1)$$

where ρ is the average receive power, $\mathbf{P} \in \mathbb{C}^{N_t \times N_s}$ ($\mathbf{W} \in \mathbb{C}^{N_r \times N_s}$) is the hybrid precoder (combiner), $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ denotes the channel matrix, $\mathbf{s} \in \mathbb{C}^{N_s \times 1}$ is the source signal vector, and $\mathbf{n} \in \mathbb{C}^{N_r \times 1}$ is the additive white Gaussian noise (AWGN) vector following the distribution $\mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_{N_r})$, where σ^2 is the noise power. In the hybrid precoding structure as illustrated in figure 1, we denote $\mathbf{P}_D \in \mathbb{C}^{N_t^{RF} \times N_s}$ ($\mathbf{W}_D \in \mathbb{C}^{N_r^{RF} \times N_s}$) as the digital precoder (combiner), and $\mathbf{P}_A \in \mathbb{C}^{N_t \times N_t^{RF}}$ ($\mathbf{W}_A \in \mathbb{C}^{N_r \times N_r^{RF}}$) as the analog precoder (combiner), respectively [13, 14, 16, 17]. To meet the constraint of transmit power, we have $\mathrm{tr}\{\mathbf{P}\mathbf{P}^H\} \le N_s$ [14]. Note that the analog precoder/combiner is realized through phase shifters [7]. Thus, all elements in $\mathbf{P}_A$ and $\mathbf{W}_A$ should have the same amplitude:

$$|\{\mathbf{P}_A\}_{i,j}| = \frac{1}{\sqrt{N_t}}, |\{\mathbf{W}_A\}_{i,j}| = \frac{1}{\sqrt{N_r}}, \quad (2)$$

where $|\cdot|$ denotes the modulus of a complex number.

2.2 Channel model

For the mmWave MIMO channel, we adopt the widely used Saleh-Valenzuela (SV) channel model for mmWave communications [8,15,16], where the channel matrix $\mathbf{H}$ is

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{L}} \left(\beta_0 \mathbf{a}_r(\varphi_0^r) \mathbf{a}_t^H(\varphi_0^t) + \sum_{i=1}^{L-1} \beta_i \mathbf{a}_r(\varphi_i^r) \mathbf{a}_t^H(\varphi_i^t) \right), \quad (3)$$

where $\beta_0 \mathbf{a}_r(\varphi_0^r) \mathbf{a}_t(\varphi_0^t)$ is the line-of-sight (LoS) component with β_0 presenting the complex gain, φ_0^r presenting the angle of arrival (AoA) at the user, φ_0^t presenting the angle of departure (AoD) at the BS, and $\beta_i \mathbf{a}_r(\varphi_i^r) \mathbf{a}_t(\varphi_i^t)$ denotes the ith non-line-of-sight (NLoS) component. In addition, \$L\$ is the total number of paths, $\mathbf{a}_r(\varphi_i^r)$ and $\mathbf{a}_t(\varphi_i^t)$ denote the array response vectors at the user and the BS, respectively. For the widely used uniform linear line antenna array (ULA) with N elements, the array response vector is [8]

$$\mathbf{a}_{ULA}(\varphi) = \sqrt{\frac{1}{N}} \left[1, e^{j\frac{2\pi}{\lambda} d \sin(\varphi)}, \cdots, e^{j(N-1)\frac{2\pi}{\lambda} d \sin(\varphi)} \right]^T, \quad (4)$$

where λ denotes the wavelength, and d is the antenna spacing. Due to limited scattering characteristics in mmWave propagation, the rank of the channel matrix $\mathbf{H}$ is much smaller than the number of antennas, i.e., the number of effective independent data streams N_s that can be exploited is limited. Therefore, we can leverage only a small number of RF chains to achieve the near-optimal performance in the hybrid precoding structure [9].

III. Proposed GMD-Based Hybrid Precoding

In this section, we first briefly review the fully digital precoding. Then, we present the proposed GMD-based hybrid precoding in detail.

3.1 Fully digital SVD- and GMD-based precoding

The SVD of the channel matrix $\mathbf{H}$ can be denoted by

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^H \\ \mathbf{V}_2^H \end{bmatrix} = \mathbf{U}_1 \mathbf{\Sigma}_1 \mathbf{V}_1^H + \mathbf{U}_2 \mathbf{\Sigma}_2 \mathbf{V}_2^H, \quad (5)$$

where $\mathbf{U}_1 \in \mathbb{C}^{N_r \times N_s}$ and $\mathbf{V}_1 \in \mathbb{C}^{N_t \times N_s}$ are semi-unitary matrices containing the left N_s columns of unitary matrices $\mathbf{U} \in \mathbb{C}^{N_r \times N_r}$ and $\mathbf{V} \in \mathbb{C}^{N_t \times N_t}$, respectively, and $\mathbf{\Sigma}_1$ is an $N_s \times N_s$ diagonal matrix with the largest N_s singular values $\sigma_1, \cdots, \sigma_{N_s}$ on its diagonal. We assume that the singular values are arranged in the decreasing order. With $\mathbf{P} = \mathbf{V}_1$ and $\mathbf{W} = \mathbf{U}_1$, the MIMO channel can be converted into N_s parallel sub-channels, where the sub-channel gains are $\sigma_1, \cdots, \sigma_{N_s}$, i.e.,

$$\mathbf{y} = \sqrt{\rho}\mathbf{W}^H\mathbf{H}\mathbf{P}\mathbf{s} + \mathbf{W}^H\mathbf{n} = \sqrt{\rho}\mathbf{U}_1^H\mathbf{H}\mathbf{V}_1\mathbf{s} + \mathbf{U}_1^H\mathbf{n} = \sqrt{\rho}\mathbf{\Sigma}_1\mathbf{s} + \mathbf{U}_1^H\mathbf{n}. \quad (6)$$

In mmWave communications, the gain of LoS component can be about 15 dB higher than that of NLoS component [9]. Consequently, the singular values of channel matrix $\mathbf{H}$ vary a lot, which results in significantly different SNRs over different sub-channels as shown in figure 2 (a). It should be pointed out

that the water-filling power allocation will further aggravate the variations of the sub-channel gains, since it allocates more power on the sub-channel with higher channel gain. If the same MCS is adopted by all sub-channels, the bit error rate (BER) performance will be primarily determined by the sub-channel with the lowest

SNR, which is unexpected in practical systems. If the similar BER performance in all sub-channels is expected, careful bit allocation is required (i.e., allocating different MCSs on different sub-channels), which will incur high coding/decoding complexity [12].

To avoid the complicated bit allocations, unlike SVD as shown in (5), we use GMD to decompose the channel [19]:

$$\mathbf{H}=\mathbf{GRQ}^{H}=\begin{bmatrix}\mathbf{G}_1 & \mathbf{G}_2\end{bmatrix}\begin{bmatrix}\mathbf{R}_1 & * \\ \mathbf{0} & \mathbf{R}_2\end{bmatrix}\begin{bmatrix}\mathbf{Q}_1^{H} \\ \mathbf{Q}_2^{H}\end{bmatrix}, \quad (7)$$

where $\mathbf{G}_1 \in \mathbb{C}^{N_r \times N_s}$ and $\mathbf{Q}_1 \in \mathbb{C}^{N_t \times N_s}$ are semi-unitary matrices containing the left N_s columns of unitary matrices $\mathbf{G} \in \mathbb{C}^{N_r \times N_r}$ and $\mathbf{Q} \in \mathbb{C}^{N_t \times N_t}$, respectively, $*$ is an arbitrary matrix, and $\mathbf{R}_1$ is an $N_s \times N_s$ upper triangular matrix with identical diagonal elements presenting the geometric mean of the largest N_s singular values $r_{i,i}=\sqrt[1/N_s]{\sigma_1\sigma_2\cdots\sigma_{N_s}}=\bar{r}, \forall i$, where $r_{i,j}$ denotes the element of $\mathbf{R}_1$ in the ith row and jth column for simplicity. Employing $\mathbf{Q}_1$ as the precoder and $\mathbf{G}_1^{H}$ as the combiner, we can rewrite (1) as

$$\begin{aligned}\mathbf{y}&=\sqrt{\rho}\mathbf{W}^{H}\mathbf{Hx}+\mathbf{W}^{H}\mathbf{n}\\&=\sqrt{\rho}\mathbf{G}_1^{H}\mathbf{HQ}_1\mathbf{s}+\mathbf{G}_1^{H}\mathbf{n}=\sqrt{\rho}\mathbf{R}_1\mathbf{s}+\mathbf{G}_1^{H}\mathbf{n}.\end{aligned} \quad (8)$$

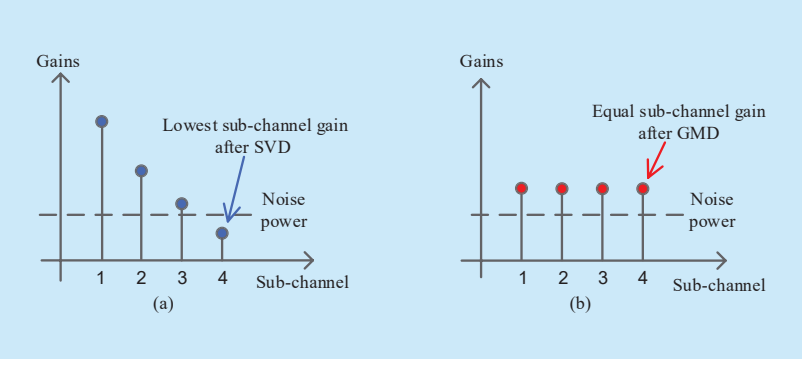


Fig. 2. *Intuitive illustrations of sub-channel gains: (a) SVD-based precoding; (b) GMD-based precoding.*

Since the equivalent channel after precoding and combining in the GMD-based precoding is an upper triangular matrix $\mathbf{R}_1$, we can utilize the successive interference cancellation at the receiver [19] to obtain N_s sub-channels with equal sub-channel gain r_{ii} as shown in figure 2 (b). Thus, we can naturally avoid the complicated bit allocation caused by different SNRs for different sub-channels in the existing SVD-based precoding. However, due to the large number of required RF chains, the energy consumption of fully digital GMD-based precoding is still high. This problem can be resolved by the proposed GMD-based hybrid precoding in the next subsection.

3.2 GMD-based hybrid precoding

In this subsection, we propose the GMD-based hybrid precoding avoid the complicated bit allocation required by the conventional SVD-based hybrid precoding. Note that we focus on the design of hybrid precoders, while the design of hybrid combiners can be realized in a similar manner. We use $\mathbf{Q}_1$, $\mathbf{Q}_A$, and $\mathbf{Q}_D$ to denote the unconstrained GMD-based precoder, the GMD-based analog precoder, and the GMD-based digital precoder, respectively. To obtain near-optimal GMD-based hybrid precoders, we should seek a pair of analog and digital precoders that are sufficiently close to the unconstraint GMD precoder. Meanwhile, the constraint on the analog precoder as shown in (2) should be also considered. In addition, the total transmit power can be bounded as $\mathrm{tr}\left(\mathbf{Q}_A\mathbf{Q}_D\mathbf{Q}_D^{H}\mathbf{Q}_A^{H}\right)\le N_s$. To sum up, the GMD-based hybrid precoder design problem can be formulated as

$$\begin{aligned}&\left(\mathbf{Q}_A^{\mathrm{opt}},\mathbf{Q}_D^{\mathrm{opt}}\right)=\arg\min_{\mathbf{Q}_A,\mathbf{Q}_D}\|\mathbf{Q}_1-\mathbf{Q}_A\mathbf{Q}_D\|_F,\\&\text{s.t. } |\{\mathbf{Q}_A\}_{i,j}|=\frac{1}{\sqrt{N_t}},\ \mathrm{tr}\left(\mathbf{Q}_A\mathbf{Q}_D\mathbf{Q}_D^{H}\mathbf{Q}_A^{H}\right)\le N_s.\end{aligned} \quad (9)$$

It is worth pointing that this norm minimization problem formulation (9) can also be interpreted as a mutual information maximization problem formulation for GMD-based hybrid precoding based on similar approxima-

tions in [8].

However, solving the optimization problem (9) is challenging, since $\mathbf{Q}_{\mathrm{A}}$ and $\mathbf{Q}_{\mathrm{D}}$ are coupled, and the constraint $|\{\mathbf{Q}_{\mathrm{A}}\}_{i,j}|=1/\sqrt{N_{\mathrm{t}}}$ is non-convex [8]. To this end, we fix $\mathbf{Q}_{\mathrm{D}}$ while designing $\mathbf{Q}_{\mathrm{A}}$, and vice versa. Note that the main challenge lies in the design of the analog precoder $\mathbf{Q}_{\mathrm{A}}$ due to the non-convex constraint. To effectively design the analog precoder $\mathbf{Q}_{\mathrm{A}}$, we first introduce the following **Lemma 1**.

Lemma 1. *Let* $\mathbf{A}_{\mathrm{t}}=\left[\mathbf{a}_{\mathrm{t}}(\phi_1^t),\mathbf{a}_{\mathrm{t}}(\phi_2^t),\cdots,\mathbf{a}_{\mathrm{t}}(\phi_L^t)\right]$ *is an* $N_t\times L$ *matrix containing all* L *steering vectors at the BS side. Then, the columns of* $\mathbf{A}_{\mathrm{t}}$ *is able to span the column space of* $\mathbf{Q}_1$.

Proof: We observe that the columns of $\mathbf{V}_1$ form an orthogonal basis of the row space for $\mathbf{H}$, while the columns of $\mathbf{A}_{\mathrm{t}}$ also span the row space of $\mathbf{H}$ according to (). Consequently, the columns of $\mathbf{A}_{\mathrm{t}}$ can span the column space of $\mathbf{V}_1$. Additionally, to implement GMD based on SVD, a reduce and conquer method is used in [19], which successively adjusts the diagonal element in $\mathbf{\Sigma}$ to $\bar{r}$ through permutations and Givens transformations. Such GMD implementation procedure (which is called "GMD procedure" in this letter) can be expressed as

$$\mathbf{Q}_1=\mathbf{V}_1\mathbf{S}_{\mathrm{R}};\ \mathbf{G}_1=\mathbf{U}_1\mathbf{S}_{\mathrm{L}};\ \mathbf{R}_1=\mathbf{S}_{\mathrm{L}}^T\mathbf{\Sigma}_1\mathbf{S}_{\mathrm{R}},\quad (10)$$

where $\mathbf{S}_{\mathrm{R}}$ and $\mathbf{S}_{\mathrm{L}}$ are $N_{\mathrm{s}}\times N_{\mathrm{s}}$ unitary matrices, which is the product of a series of permutation matrices and Givens matrices [19]. From (10), we observe that if the columns of an arbitrary matrix $\mathbf{\Omega}$ spans the column space of $\mathbf{V}_1$, they can also span the column space of $\mathbf{Q}_1$, since multiplying by $\mathbf{S}_{\mathrm{R}}$ does not change the column space of $\mathbf{V}_1$. Therefore, the columns of $\mathbf{A}_{\mathrm{t}}$ can span the column space of $\mathbf{Q}_1$.

Recalling the objective function in (9), we find that the form of hybrid precoders $\mathbf{Q}_{\mathrm{A}}\mathbf{Q}_{\mathrm{D}}$ can also be interpreted as treating the analog precoder $\mathbf{Q}_{\mathrm{A}}$ as a basis and using the digital precoder $\mathbf{Q}_{\mathrm{D}}$ as the combination coefficients. Therefore, based on the **Lemma 1**, it is natural to match up these two ``basis'', i.e., leveraging the $\mathbf{A}_{\mathrm{t}}$ to serve as $\mathbf{Q}_{\mathrm{A}}$. Besides, $\mathbf{A}_{\mathrm{t}}$ also satisfies the constant modulus constraints. However, rigorously speaking, $\mathbf{A}_{\mathrm{t}}$ cannot be directly utilized as the analog precoder $\mathbf{Q}_{\mathrm{A}}$, since $\mathbf{A}_{\mathrm{t}}$ has L columns while $\mathbf{Q}_{\mathrm{A}}$ has $N_{\mathrm{t}}^{\mathrm{RF}}$ columns. Note that in hybrid precoding structure, $N_{\mathrm{t}}^{\mathrm{RF}}$ is usually assumed to be smaller than L [8, 16]. Therefore, we need to select the ``best'' $N_{\mathrm{t}}^{\mathrm{RF}}$ columns from $\mathbf{A}_{\mathrm{t}}$ to form $\mathbf{Q}_{\mathrm{A}}$. In this way, the analog precoder design problem becomes

$$\begin{aligned}\mathbf{T}&=\arg\min_{\mathbf{T}}\|\mathbf{Q}_1-\mathbf{A}_{\mathrm{t}}\mathbf{T}\mathbf{Q}_{\mathrm{D}}\|_F,\\ &\text{s.t.}\ \|\mathrm{diag}(\mathbf{T}\mathbf{T}^H)\|_0=N_{\mathrm{t}}^{\mathrm{RF}},\end{aligned}\quad (11)$$

where $\mathbf{T}$ is a selecting matrix with only $N_{\mathrm{t}}^{\mathrm{RF}}$ non-zero rows. Since the number of RF chains is smaller than the number of paths in hybrid precoding, and we only pick up the steering vectors with the highest correlations to the optimal precoder $\mathbf{Q}_1$, $\mathbf{T}$ will be a sparse matrix, where only a small number of elements is non-zero. Thus, we can leverage the sparse-aware algorithms, such as the orthogonal matching pursuit (OMP) algorithm [8] to solve (11). Note that we remove the power constraint on the precoders in (11), since it can be fulfilled through normalization on the digital precoder.

After the analog precoder is determined, the digital precoder design problem becomes a Frobenius norm minimization problem:

$$\begin{aligned}\mathbf{Q}_{\mathrm{D}}^{\mathrm{opt}}&=\arg\min_{\mathbf{Q}_{\mathrm{D}}}\|\mathbf{Q}_1-\mathbf{Q}_{\mathrm{A}}\mathbf{Q}_{\mathrm{D}}\|_F,\\ &\text{s.t.}\quad \mathrm{tr}\left(\mathbf{Q}_{\mathrm{A}}\mathbf{Q}_{\mathrm{D}}\mathbf{Q}_{\mathrm{D}}^H\mathbf{Q}_{\mathrm{A}}^H\right)\le N_{\mathrm{s}},\end{aligned}\quad (12)$$

where $\mathbf{Q}_{\mathrm{A}}$ is the determined analog precoder obtained by solving (11). The optimal solution to (12) has a least square form [16]. So the digital precoder can be designed as

$$\mathbf{Q}_{\mathrm{D}}=\{\mathbf{Q}_{\mathrm{A}}\}^{\dagger}\mathbf{Q}_1,\quad (13)$$

and the power constraint in () can be satisfied through normalizing $\mathbf{Q}_{\mathrm{D}}$.

The proposed GMD-based hybrid precoding algorithm has been summarized in **Algorithm 1**, which is mainly composed of two parts. The first part including steps 4-9 corresponds to the construction of the analog precoder $\mathbf{Q}_{\mathrm{A}}$,

Algorithm 1. The proposed GMD-based hybrid precoding algorithm.

Input: Channel matrix $\mathbf{H}$

Output: Precoding matrix $\mathbf{Q}_{\mathrm{A}}$ and $\mathbf{Q}_{\mathrm{D}}$

1. Construct $\mathbf{A}_{\mathrm{t}}$, perform GMD of the channel matrix: $[\mathbf{G}_1, \mathbf{R}_1, \mathbf{Q}_1] = \mathrm{GMD}(\mathbf{H})$;
2. Initialize the residual matrix $\mathbf{Q}_{\mathrm{res}} = \mathbf{V}_1$, and the index set $\mathcal{A}_{\mathrm{t}} = \varnothing$;
3. **Repeat**:
4. $j = \mathrm{argmax}_{l=1,\cdots,L} (\mathbf{A}_{\mathrm{t}}^H \mathbf{Q}_{\mathrm{res}} \mathbf{Q}_{\mathrm{res}}^H \mathbf{A}_{\mathrm{t}})_{l,l}$;
5. $\mathcal{A}_{\mathrm{t}} = \mathcal{A}_{\mathrm{t}} \cup j$, and $\mathbf{\Xi} = \{\mathbf{a}_{\mathrm{t}}(\phi_j^{\mathrm{t}})\}$, $j \in \mathcal{A}_{\mathrm{t}}$;
6. $\mathbf{Y} = \mathbf{\Xi}^{\dagger} \mathbf{Q}_1$;
7. $\mathbf{Q}_{\mathrm{res}} = \dfrac{\mathbf{Q}_1 - \mathbf{\Xi}\mathbf{Y}}{\| \mathbf{Q}_1 - \mathbf{\Xi}\mathbf{Y} \|_F}$;
8. **Until** ($| \mathcal{A}_{\mathrm{t}} | == N_{\mathrm{t}}^{\mathrm{RF}}$);
9. Construct $\mathbf{Q}_{\mathrm{A}} = \mathbf{\Xi}$;
10. Construct $\mathbf{Q}_{\mathrm{D}} = \{\mathbf{Q}_{\mathrm{A}}\}^{\dagger} \mathbf{Q}_1$;
11. Normalize $\mathbf{Q}_{\mathrm{D}} = \sqrt{N_{\mathrm{s}}} \dfrac{\mathbf{Q}_{\mathrm{D}}}{\| \mathbf{Q}_{\mathrm{A}} \mathbf{Q}_{\mathrm{D}} \|_F}$;

where we use the OMP algorithm to greedily select the $N_{\mathrm{t}}^{\mathrm{RF}}$ BS steering vectors and store them in $\mathbf{\Xi}$. The second part including steps 10-11 is the construction and normalization of the digital precoder $\mathbf{Q}_{\mathrm{D}}$ based on (13).

Finally, we analyze the complexity of the proposed GMD-based hybrid precoding, which is composed of three parts in general. The first part is implementing GMD based on SVD, whose complexity is $\mathcal{O}((N_{\mathrm{s}} + N_{\mathrm{t}}) N_{\mathrm{s}})$ [19], since only an $N_{\mathrm{s}} \times N_{\mathrm{s}}$ matrix $\mathbf{S}_{\mathrm{R}}$ is multiplied on $\mathbf{V}_1$. The second part is the construction of the analog precoder based on the OMP algorithm, whose complexity is $\mathcal{O}\left((N_{\mathrm{t}}^{\mathrm{RF}})^2 N_{\mathrm{t}} N_{\mathrm{s}}\right)$. The third part is the computation of the digital precoder, of which the complexity is $\mathcal{O}(N_{\mathrm{t}}^{\mathrm{RF}} N_{\mathrm{t}} N_{\mathrm{s}})$, as we need to compute the pseudo-inverse matrix of $\mathbf{\Xi}$. Hence, the overall complexity of GMD-based hybrid precoding is $\mathcal{O}\left(N_{\mathrm{t}} N_{\mathrm{s}} \left((N_{\mathrm{t}}^{\mathrm{RF}})^2 + N_{\mathrm{t}}^{\mathrm{RF}} + 1\right) + N_{\mathrm{s}}^2\right)$, which is comparable to the conventional SVD-based precoding [8, 10].

IV. Simulation Results

In this section, we evaluate the performance of the proposed GMD-based hybrid precoding through simulations. We consider a typical mmWave massive MIMO system at 28 GHz, where an $N_{\mathrm{t}} = 256$-element ULA with the antenna spacing $d = \lambda / 2$ is employed at the BS, while an $N_{\mathrm{r}} = 16$-element ULA with also $d = \lambda / 2$ is employed at the user [8, 15]. Both BS and the user adopt $N_{\mathrm{t}}^{\mathrm{RF}} = N_{\mathrm{r}}^{\mathrm{RF}} = 4$ RF chains. For the mmWave channel, there are one LoS path and 4 NLoS paths. Additionally, the LoS path gain β_0 follows the distribution $\mathcal{CN}(0,1)$, while the NLoS path gains β_i follow the distribution $\mathcal{CN}(0,10^{-\mu})$, where μ is a power normalizing factor to adjust the power distribution between LoS component and NLoS component [8, 15]. The AoA and AoD of each component are uniformly distributed in $\left[-\pi/2, \pi/2\right]$ [8, 15]. For the modulation scheme, we adopt the 16QAM modulation on all sub-channels after SVD/GMD to guarantee the same complexity and compare the performance of different precoding methods.

4.1 Performance comparison under perfect CSI

In this subsection, we evaluate the performance of the proposed GMD-based hybrid precoding with the assumption that perfect channel state information (CSI) is available at both the transmitter and the receiver.

Firstly, we consider an LoS environment in the mmWave propagations [15]. Figure 3 shows the results of BER performance comparison, where $\mu = 1.5$ (the power of LoS component is 15 dB higher that of NLoS component). We can observe that the proposed GMD-based precoding (including the fully digital GMD-based precoding and the GMD-based hybrid precoding) can achieve better BER performance than the conventional SVD-based precoding (including the fully digital SVD-based precoding and the SVD-based hybrid precoding [8]). Since the GMD-based precoding converts the mmWave massive MIMO channel into several sub-channels with identical SNR, we can naturally avoid the severe BER performance degradation in sub-channels with very low SNRs in SVD-based precoding. Furthermore, the perfor-

mance gap between the fully digital GMD-based precoding and the proposed GMD-based hybrid precoding is negligible, which implies that the proposed GMD-based hybrid precoding is able to approach the performance of the fully digital GMD-based precoding.

We also consider the NLoS environment in the performance comparison [18], where μ is set to zero to indicate the same power of the LoS paht and the NLoS paths. The result is given in figure 4, where we find that the approximation performance of hybrid precoding to fully digital precoding degrades, since the power of the channel disperses onto several paths which cannot be collected via a limited number of RF chains. In addition, we also observe that the performance gap between the GMD-based precoding and SVD-based precoding becomes smaller, because the gains of sub-channels after precoding and combining tend to be more similar in an NLoS environment. Nevertheless, the proposed GMD-based hybrid precoding can still outperform the conventional precoding schemes.

4.2 Impacts of low-rank CSI

In practical mmWave massive MIMO systems, acquiring perfect CSI is very challenging due to the high dimension of the channel matrix. Therefore, it is important to evaluate the performance of the proposed GMD-based hybrid precoding under imperfect CSI. In mmWave MIMO systems, a low-rank channel estimation is usually conducted instead of a full-rank channel estimation to save the signaling overhead [21], where only a limited number of paths are estimated (e.g., estimating the 4 paths with the highest power instead of the whole 10 paths when $N_{\mathrm{RF}}=4$ to enable the hybrid precoding). For the channel estimation schemes, we adopt a hierarchical codebook [21, 22] to estimate the AoA, AoD, and complex gain for a path through analog beam training, and the number of estimated paths is equal to the number of RF chains.

Figure 5 presents the BER performance comparison in an LoS environment where

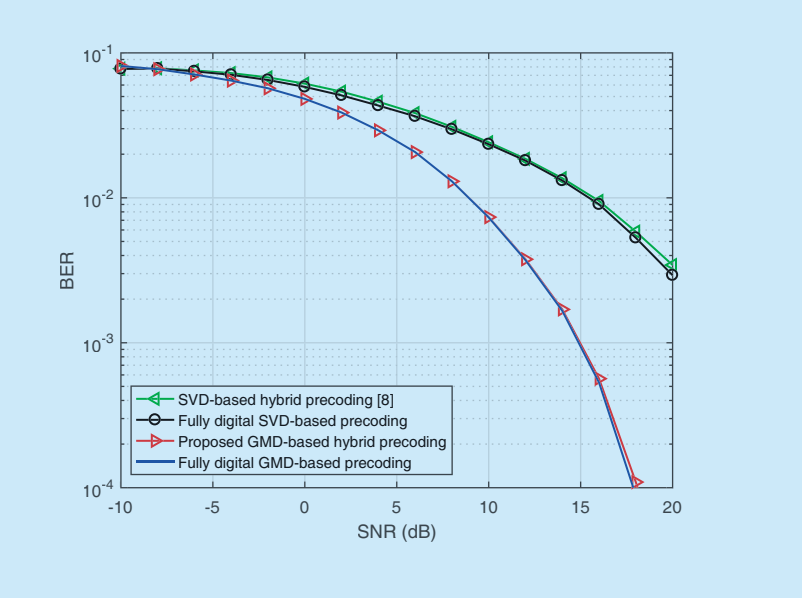


Fig. 3. *BER performance comparison in the LoS environment with* $\mu=1.5$.

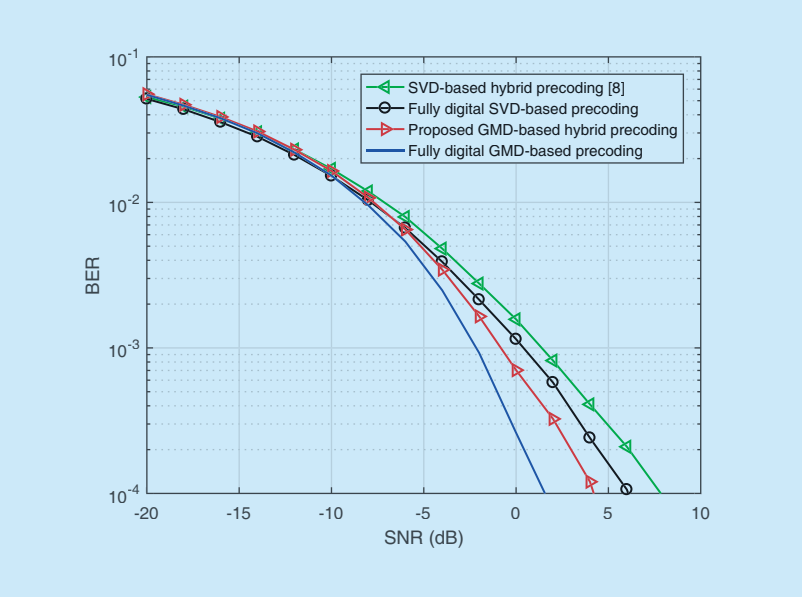


Fig. 4. *BER performance comparison in the NLoS environment with* $\mu=0$.

$\mu=1.5$. From the figure, we can observe that using the low-rank CSI to precode the signals will suffer from an obvious performance loss when the channel presents an LoS characteristic. This is because omitting the weak paths in channel estimation will make the singular values corresponding to the weak paths much smaller, which obviously lowers the geometric mean value of these singular values, i.e.,

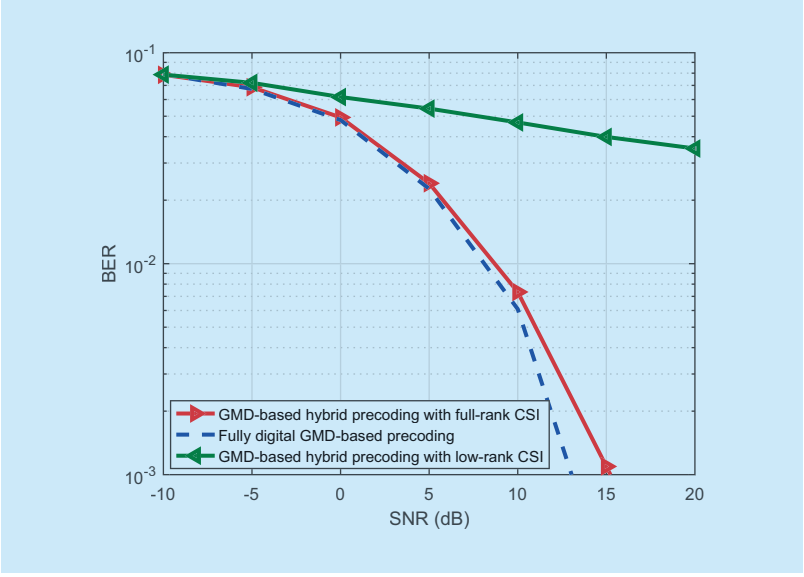


Fig. 5. *BER performance comparison between precoding with full-rank CSI and precoding with low-rank CSI in the LoS environment.*

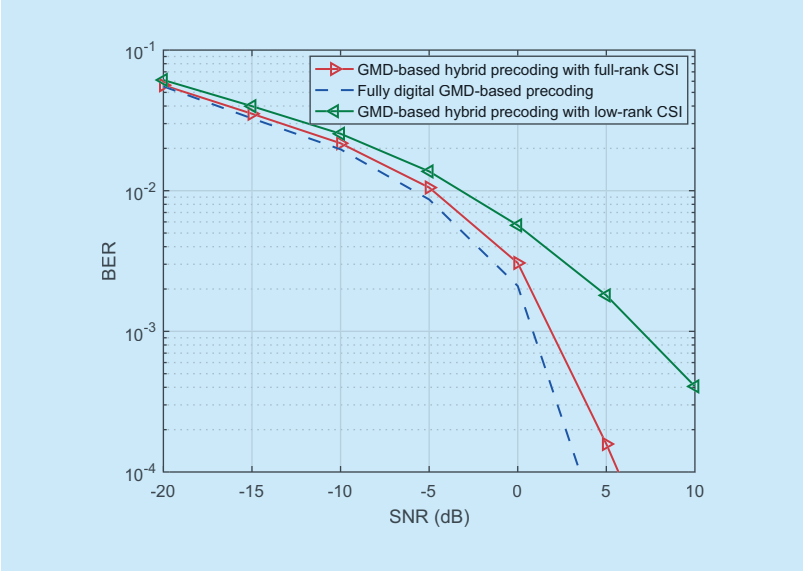


Fig. 6. *BER performance comparison between full-rank CSI and precoding with low-rank CSI in the NLoS environment.*

the gain of each sub-channel after GMD. In addition, the performance of the analog beam training also degrades when the estimating the NLoS paths with very small power in the LoS environment, which further aggravates the performance loss.

We also evaluate the impact of channel estimation in the NLoS environment in figure 6, where $\mu = 0$. In this figure, we find that the performance gap between precoding with full-rank CSI and precoding with low-rank CSI becomes smaller. On the one hand, the singular values of the channel matrix are more robust when there are several paths with similar power. On the other hand, the performance of the beam training method becomes better when estimating the paths with relative higher power.

V. Conclusions

In this paper, a GMD-based hybrid precoding is proposed to avoid the complicated bit allocation in the conventional SVD-based hybrid precoding. We have found that with the help of GMD, the mmWave MIMO channel can be converted into several sub-channels with identical SNRs, thus naturally avoiding the complicated bit allocation. Furthermore, we have proposed to decouple the design of the analog and digital precoders, where the analog precoder is designed using the principle of basis pursuit, while the digital precoder is obtained by using GMD. Simulation results verify that the proposed GMD-based hybrid precoding can achieve better performance than the conventional SVD-based hybrid precoding.

Acknowledgement

This work was supported by the National Natural Science Foundation of China for Outstanding Young Scholars (Grant No. 61722109), the National Natural Science Foundation of China (Grant No. 61571270), and the Royal Academy of Engineering through the UK–China Industry Academia Partnership Programme Scheme (Grant No. UK-CIAPP\49).

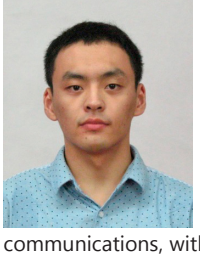
References

[1] S. Mumtaz, J. Rodriquez, and L. Dai, *MmWave Massive MIMO: A Paradigm for 5G*, Academic Press, Elsevier, 2016.

[2] W. Feng, Y. Wang, D. Lin, N. Ge, J. Lu, and S. Li, "When mmWave communications meet network densification: A scalable interference coordination perspective," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, Jul. 2017, pp. 1459–1471.

[3] B. Wang, L. Dai, Z. Wang, N. Ge, and S. Zhou, "Spectrum and energy efficient beamspace MIMO-NOMA for millimeter-wave communications using lens antenna array," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, Oct. 2017, pp. 2370-2382.
[4] X. Li, T. Jiang, S. Cui, J. An, and Q. Zhang, "Cooperative communications based on rateless network coding in distributed MIMO systems," *IEEE Wireless Commun.*, vol. 17, no. 3, Jun. 2010, pp. 60–67.
[5] Z. Xiao, P. Xia, and X.-G. Xia, "Enabling UAV cellular with millimeterWave communication: Potentials and approaches," *IEEE Commun. Mag.*, vol. 54, no. 5, May 2016, pp. 66-73.
[6] J. Zhang, Y. Zhang, Y. Yu, R. Xu, Q. Zheng, P. Zhang, "3D MIMO: How much does it meet our expectation bbserved from channel measurements?" *IEEE J. Sel. Areas Commun.*, vol. 35, no. 8, Aug. 2017, pp. 1887–1903.
[7] S. Han, C.-L. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid precoding analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, Jan. 2015, pp. 186–194.
[8] El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, Mar. 2014, pp. 1499–1513.
[9] Z. Gao, L. Dai, D. Mi, Z. Wang, M. A. Imran, and M. Z. Shakir, "Mmwave massive-MIMO-based wireless backhaul for the 5G ultra-dense network," *IEEE Wireless Commun.*, vol. 22, no. 5, Oct. 2015, pp. 13–21.
[10] Z. Xu, S. Han, Z. Pan, and C.-L. I, "Alternating beamforming methods for hybrid analog and digital MIMO transmission," *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 1595–1600.
[11] R. Zi, X. Ge, J. Thompson, C.-X. Wang, H. Wang, and T. Han, "Energy efficiency optimization of 5G radio frequency chain systems," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, Apr. 2016, pp. 758–771.
[12] C.-L. Chao, S.-H. Tsai, and T.-Y. Hsu, "Bit allocation schemes for MIMO equal gain precoding," *IEEE Trans. Wireless Commun.*, vol. 10, no. 5, Mar. 2011, pp. 1345–1350.
[13] Z. Xiao, X.-G. Xia, D. Jin, and N. Ge, "Iterative eigenvalue decomposition and multipath-grouping Tx/Rx joint beamformings for millimeterwave communications," *IEEE Tran. Wireless Commun.*, vol. 14, no. 3, Mar. 2015, pp. 1595–1607.
[14] X. Xue, Y. Wang, X. Wang, and T. E. Bogale, "Joint source and relay precoding in multi-antenna millimeter-wave systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, Oct. 2016, pp. 4924–4937.
[15] X. Gao, L. Dai, S. Han, C.-L. I, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, Apr. 2016, pp. 998–1009.
[16] S. He, C. Qi, Y. Wu, and Y. Huang, "Energy-efficient transceiver design for hybrid sub-array architecture MIMO systems," *IEEE Access*, vol. 4, Jan. 2017, pp. 9895–9905.
[17] S. Han, C.-L. I, Z. Xu, and S. Wang, "Reference Signals Design for Hybrid Analog and Digital Beamforming," *IEEE Commun. Lett.*, vol. 18, no. 7, Jul. 2014, pp. 1191–1193.
[18] Z. Xiao, P. Xia, X.-G. Xia, "Channel estimation and hybrid precoding for millimeter-wave MIMO systems: A low-complexity overall solution", *IEEE Access*, vol. 5, Jul. 2017, pp. 16100–16110.
[19] C.-E. Chen, Y.-C. Tsai, and C.-H. Yang, "An iterative geometric mean decomposition algorithm for MIMO communications systems," IEEE *Tran. Wireless Commun.*, vol. 14, no. 1, Jan. 2015, pp. 343–352.
[20] S. He, J. Wang, Y. Huang, B. Otterste, W. Hong, "Codebook-based hybrid precoding for millimeter wave multiuser systems," *IEEE Trans. Signal Process.*, vol. 64, no. 20, Oct. 2017, pp. 5289–5304.
[21] Z. Xiao, P. Xia, and X.-G. Xia, "Codebook design for millimeterwave channel estimation with hybrid precoding structure", *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, Jan. 2017, pp. 141–153.
[22] Z. Xiao, T. He, P. Xia, and X.-G. Xia, "Hierarchical codebook design for beamforming training in millimeter-wave communication," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, May 2016, pp. 3380–3392.

Biographies



Tian Xie, received the B.E. degree from Tsinghua University, Beijing, China, in 2015. He is currently working towards Master degree in Electronic Engineering of Tsinghua University, Beijing, China. His research interests are in wireless communications, with a focus on Massive MIMO and mmWave communications.

Linglong Dai, received the Ph.D. degree (with the highest honor) from Tsinghua University, Beijing, China, in 2011. From 2011 to 2013, he was a Postdoctoral Research Fellow with the Department of Electronic Engineering, Tsinghua University, where he has been an Assistant Professor since July 2013 and then an Associate Professor since June 2016. His current research interests include massive

MIMO, millimeter-wave communications, multiple access, and sparse signal processing.

Xinyu Gao, received the B.E. degree of Communication Engineering from Harbin Institute of Technology, Heilongjiang, China in 2014. He is currently working towards Ph. D. degree in Electronic Engineering from Tsinghua University, Beijing, China. His research interests include massive MIMO and mmWave communications, with the emphasis on signal detection and precoding.

Muhammad Zeeshan Shakir, received his Ph.D. degree in electronic and electrical engineering from the University of Strathclyde, Glasgow, UK, in 2010. Since 2016, he has been with the School of Engineering and Computing at University of the West Scotland (UWS), Glasgow, Scotland. His research interests include performance analysis of wireless communication systems, such as heterogeneous small-cell networks and cooperative and cognitive communications networks.

Jianjun Li, received the Ph.D. degree in electronic engineering from Tsinghua University, Beijing, China, in 2002. From 2007~2015, he has worked in Posdata, Pantech and Innovative technology Lab. Co. as a senior researcher in Korea. Since 2015, he has been a Professor at Zhongyuan University of Technology and visiting scholar in Tsinghua University. His current research interest is on 5G wireless communication including massive MIMO, NOMA, and new waveform designs.