

Received January 2, 2019, accepted January 9, 2019, date of publication January 22, 2019, date of current version February 8, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2894122

# Dimension-Deficient Channel Estimation of Hybrid Beamforming Based on Compressive Sensing

YU XIAO<sup>1</sup>, (Student Member, IEEE), YAFENG WANG<sup>1</sup>, (Senior Member, IEEE),  
AND WEI XIANG<sup>2</sup>, (Senior Member, IEEE)

<sup>1</sup>Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China

<sup>2</sup>College of Science and Engineering, James Cook University, Cairns, QLD 4878, Australia

Corresponding authors: Yafeng Wang (wangyf@bupt.edu.cn) and Wei Xiang (wei.xiang@jcu.edu.au)

This work was supported by the National Key Technology Research and Development Program of China under Grant 2017ZX03001012-003.

**ABSTRACT** Due to high hardware costs for digital beamforming, hybrid beamforming (HBF) is widely employed in millimeter-wave (mmWave) communications systems. However, the number of radio frequency chains in the analog part of HBF is far less than that of antennas, which causes a serious dimension-deficient problem. In order to overcome this problem, this paper proposes a compressive sensing algorithm using an adaptive overcomplete dictionary to estimate the sparse channel in the HBF-based mmWave system. The algorithm adaptively generates the dictionary by using the received signal to accurately reconstruct the mmWave channel. The simulation results are presented to demonstrate that the proposed algorithm outperforms its traditional counterparts in terms of the normalized mean square error and the spectral efficiency.

**INDEX TERMS** Sparse channel estimation, hybrid beamforming, adaptive overcomplete dictionary, compressive sensing.

## I. INTRODUCTION

In order to meet the explosive demands of user data growth, the current spectrum scarcity crisis needs to be addressed [1]–[3]. Thanks to abundant spectral resources, mmWave communications are widely studied in the fifth generation (5G) mobile networks. Although offloading data to mmWave spectra can improve channel capacity, it is not without sacrifices. Air is a highly absorptive dielectric with respect to the mmWave spectra, so the resulting path losses are very severe, thereby limiting the coverage of a base station [4]. However, mmWave has shorter wavelengths than the current LTE band, which means it is capable of packing more antennas in a limited physical space. The increased number of antennas will bring about more channel gains through the use of techniques such as precoding. Therefore, massive multiple-input multiple-output (MIMO) technology can be fully utilized in mmWave bands.

For the purpose of overcoming the shortcomings of mmWave communications and taking advantage of the gains brought by multiple antennas, digital beamforming (DBF) technology is widely considered. However, in the number of antennas, hardware costs become the main bottleneck

with the increase in massive MIMO. Then analog beamforming (ABF) is adopted to reduce hardware costs, which evolves into HBF technology. HBF is adopted by 5G wireless communication systems to achieve better beamforming gains. Because of the use of analog components, HBF usually cannot attain the same performance of DBF. Nevertheless, it is proved in [4] that HBF is able to achieve almost the same performance as DBF in terms of SE, when the number of RF chains is twice as that of data streams or more. While the algorithm proposed in [4] focuses primarily on the single-user scenario. When extended to the scenario of multiple users, SE will decrease at low signal-to-noise ratios (SNRs). In [5], the algorithm is extended from the single-user scenario to the multi-user one, while it proposes a different HBF design method that can maximize system capacity. To address the problem of multiple users, a novel algorithm is proposed in [6]. The proposed HBF matrix for multiple users is not only superior in limited feedback and training overhead, but also helps achieve an excellent system performance. However, it is noted that the HBF performance in [6] is nearly equal to its DBF counterpart, only when data are transmitted with a single antenna.

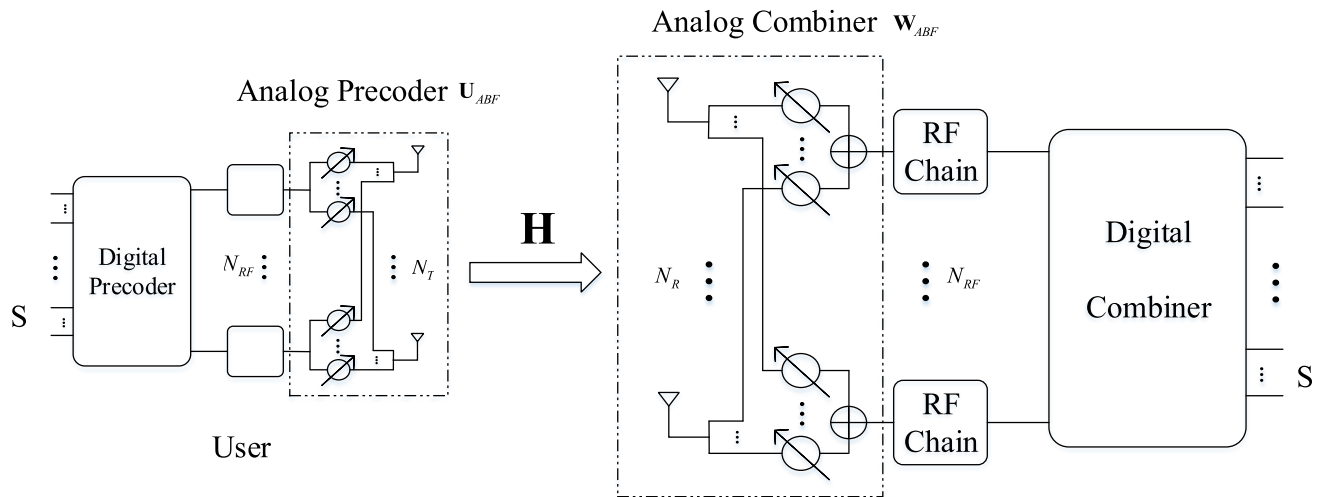


FIGURE 1. Block diagram of the mmWave communications system with full-connection HBF.

System coverage degrades with the increase in the number of channels.

To address the challenges of channel estimation (CE) in HBF, an adaptive algorithm is proposed to estimate the angle of arrival (AoA) and the angle of departure (AoD) for channel reconstruction in [7]. An algorithm is presented in lieu of the singular value decomposition (SVD) for channel estimation, which can reduce the computational complexity and the feedback overhead [8]. In [9], an asymmetric beam search scheme is used to estimate the channel taking into account the sparsity of the mmWave channel. On one hand, this scheme employs CS to reduce the training overhead at the transmitter. On the other hand, the receiver utilizes exhaustive beam training to ensure a robust performance at low SNRs.

Different from the CS-based method to estimate the AoA and AoD matrix of channel, Guo *et al.* [10] present a two-dimensional multiple signal classification (2-D MUSIC) algorithm used in the beam space to estimate the directions of the mmWave channel. It is different from the conventional MUSIC algorithm used in the element space, which can only estimate one dimension (i.e. AoA or AoD) at a time and also does not consider beamforming. While it is stated in [10] that the beamspace MUSIC algorithm suffers from the problem of spectrum ambiguity, which means the quantized ABF may cause the miss match problem during the process of finding the extreme point of the directional spectrum. The 2-D MUSIC algorithm is also studied in [11], where an efficient 2-D direction-finding MUSIC algorithm based on the double polynomial root finding procedure is proposed to estimate the AoA and AoD jointly.

In most cases, the number of RF chains is far less than that of antennas. Therefore, the received signal does not contain full channel state information (CSI). This phenomenon, which is dubbed dimension-deficient, makes channel estimation in HBF different from conventional

channel estimation. In order to tackle the dimension-deficient challenge, this paper proposes a CS-based channel estimation algorithm for uplink single-user MIMO systems. Compared with conventional channel estimation methods, the proposed algorithm designs an adaptive overcomplete dictionary to improve the reconstruction probability at low SNRs, in an effort to reduce the NMSE.

**Contribution:** this paper proposes a dictionary-adaptive compressive sensing channel estimation algorithm to overcome the dimension-deficient problem in HBF. Compared with conventional CS channel estimation methods in massive MIMO systems, the novelty of the proposed algorithm lies in the construction of an adaptive dictionary to deal with the dimension-deficient issue, improve the robustness against noise at low SNRs, minimize the NMSE of channel estimation, and achieve nearly identical SE.

*Notation:*  $\mathbf{X}$ ,  $\mathbf{x}$  and  $x$  represent a matrix, vector and scalar, respectively.  $(*)^T$ ,  $(*)^H$ ,  $(*)^{-1}$ ,  $(*)^\dagger$ ,  $vec(*)$  and  $\|*\|_F$  denote the transport, conjugate transport, inverse, pseudo-inverse, vectorize, and Frobenius norm of a matrix, respectively.  $\otimes$  indicates the Kronecker product operator.

This paper is organized as follows. Section II describes the system and mmWave channel models. Section III proposes the CS-based channel estimation scheme to cope with the dimension-deficient challenge in HBF. Section IV presents the simulation results of the proposed algorithm, and compares them with those of the conventional channel estimation algorithms in [8], [10], and [18]. Finally, concluding remarks are drawn in Section V.

## II. SYSTEM MODEL

Consider an mmWave massive MIMO system with a single UE (SU-MIMO) and gNB in 5G. It is well known that HBF has two different structures, i.e. fully connected and partially connected HBF [20], [21]. Each RF chain in full-connection

HBF is connected to every antenna, so all the antennas can be optimized jointly and the transceiver have a finer beam with full beamforming gains. On the other hand, each RF chain in partially connected HBF is connected to a subset of the antennas, so this structure has lower hardware complexity than its fully connected counterpart, while its array gain is lower and the beam width is rougher. Consequently, this paper chooses fully connected HBF as shown in Fig. 1. Due to HBF being usually implemented on both uplink and downlink sides, we will only consider the uplink channel estimation at the gNB side for simplicity without loss of generality. The UE is equipped with  $N_T$  antennas and  $N_{RF}$  RF chains, while the gNB is equipped with  $N_R$  antennas and  $N_{RF}$  RF chains. It is noted that the numbers of RF chains are generally different for the UE and gNB. But this fact has no impact on the research under consideration in this paper, so it is assumed that the numbers of RF chains at the UE and gNB are the same. The number of RF chains should be far less than that of antennas, i.e.,  $N_{RF} \ll N_R$ .

Considering the sparsity of the mmWave channel, the following channel model that has been widely considered in the literature is adopted [6]–[9], [16]

$$\mathbf{H} = \sqrt{\frac{N_T N_R}{L}} \sum_{l=1}^L \alpha_l \mathbf{a}_r(\theta_l^r) \mathbf{a}_t^H(\theta_l^t), \quad (1)$$

where  $L$  is the number of channel paths,  $\alpha_l \sim \mathcal{CN}(0, \sigma_\alpha^2)$  is the channel gain,  $\mathbf{a}_r(\theta_l^r)$  and  $\mathbf{a}_t(\theta_l^t)$  are the arrival steering and departure steering vectors, respectively.  $\theta_l^t$  and  $\theta_l^r$  are the azimuth angles of departure and arrival uniformly distributed in  $[0, 2\pi)$ . Assuming the uniform linear arrays (ULA) antennas, the steering vectors are shown as  $\mathbf{a}_r(\theta_l^r) = \frac{1}{\sqrt{N_R}} [1, e^{j\frac{2\pi}{\lambda} d \sin(\theta_l^r)}, \dots, e^{j(N_R-1)\frac{2\pi}{\lambda} d \sin(\theta_l^r)}]^T$  and  $\mathbf{a}_t(\theta_l^t) = \frac{1}{\sqrt{N_T}} [1, e^{j\frac{2\pi}{\lambda} d \sin(\theta_l^t)}, \dots, e^{j(N_T-1)\frac{2\pi}{\lambda} d \sin(\theta_l^t)}]^T$ , where  $d = \lambda/2$  and  $\lambda$  is the carrier wavelength. Eqn. (1) can be rewritten in a more compact form as follows

$$\mathbf{H} = \mathbf{A}_R \mathbf{D}_\alpha \mathbf{A}_T^H, \quad (2)$$

where  $\mathbf{A}_R = [\mathbf{a}_r(\theta_1^r), \dots, \mathbf{a}_r(\theta_L^r)]$ ,  $\mathbf{A}_T = [\mathbf{a}_t(\theta_1^t), \dots, \mathbf{a}_t(\theta_L^t)]$ ,  $\mathbf{D}_\alpha \in \mathbb{C}^{L \times L}$  is a square matrix with  $[\alpha_1, \dots, \alpha_L]$  in its diagonal line. As  $\mathbf{A}_R$  and  $\mathbf{A}_T$  is unknown in the channel estimation process, we use virtual channel representation in [17] to rewrite the channel model as

$$\mathbf{H} = \mathbf{V}_R \mathbf{H}_V \mathbf{V}_T^H, \quad (3)$$

where  $\mathbf{V}_R$  and  $\mathbf{V}_T$  are the unitary discrete Fourier transform (DFT) matrices of sizes  $N_R \times N_R$  and  $N_T \times N_T$ , respectively.  $\mathbf{H}_V$  is the virtual channel element matrix of dimension  $N_R \times N_T$ . Virtual channel representation transforms a channel model in the spatial domain into the beam-space or the wave-number domain. In other words, virtual representation describes the channel by using fixed virtual angles which are decided by the spatial resolution ( $N_R \times N_T$ ) of the arrays [17]. If it is assumed that the real channel

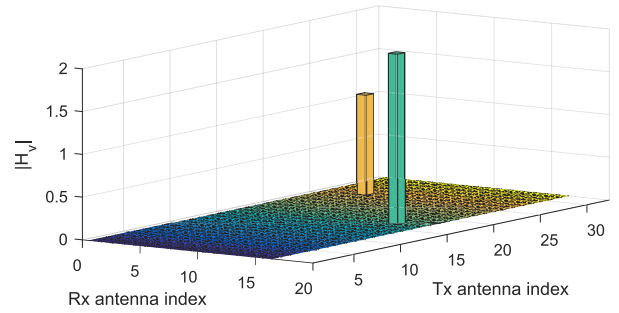


FIGURE 2. Virtual channel matrix without leakage.

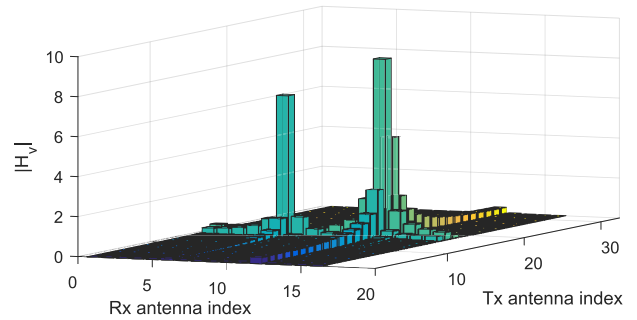


FIGURE 3. Virtual channel matrix with leakage.

matrix  $\mathbf{H}$  is known, the relationship between the virtual channel coefficient matrix  $\mathbf{H}_V$  and the real channel matrix can be written as

$$\mathbf{H}_V = \mathbf{V}_R^H \mathbf{H} \mathbf{V}_T, \quad (4)$$

since  $\mathbf{V}_R$  and  $\mathbf{V}_T$  are DFT matrices and they share property that  $\mathbf{V}_{R/T}^H \mathbf{V}_{R/T} = \mathbf{I}$ , where  $\mathbf{I}$  is the identity matrix.

It should be noted that  $\mathbf{H}_V$  suffers from leakage from a specific bin to adjacent bins, which means the number of bins with non-zero values are not  $L$ , and the adjacent bins also have relatively large amplitudes. For instance, there is a virtual channel transformed from the real channel with two channel paths ( $L = 2$  in this example) by using (4). Fig. 2 and 3 compare the difference between the virtual channels with and without adjacent leakage. Fig.2 plots the virtual channel with two paths without adjacent leakage. The magnitudes of the two main channel paths are clearly distinct from adjacent bins which have zero or very small magnitudes. However, Fig. 3 illustrates the virtual channel with adjacent leakage. Although the two main channel paths can be seen clearly, the adjacent bins also have non-negligible magnitudes. In other words, the adjacent bins may also contain part of the CSI and should be considered in channel estimation. Consequently, the sparsity of the virtual channel is generally more than the number of channel paths.

It should be noted that the adjacent leakage problem caused by virtual channel representation is different from the

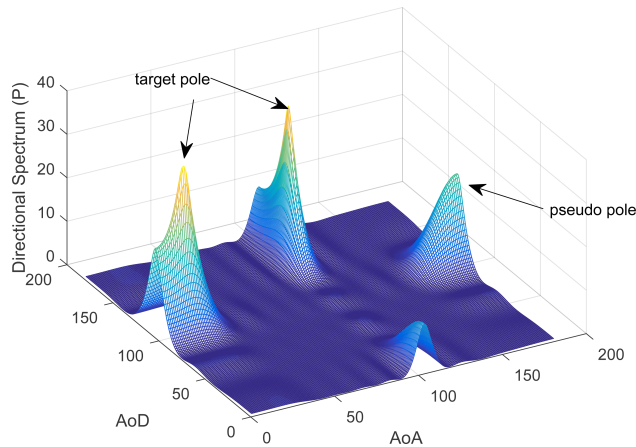


FIGURE 4. Spectrum ambiguity in 2-D MUSIC.

spectrum ambiguity problem mentioned in [10]. As shown in Fig. 4, the spectrum ambiguity problem is that the extreme point of the directional spectrum may be more than the targets we look for. So when searching for the poles of the directional function, the pseudo pole may take the place of the target pole.

For the uplink, the UE transmits the pilot signal  $\mathbf{X} \in \mathbb{C}^{N_T \times X_L}$  to the gNB, where  $X_L$  is the length of the signal transmitted over each antenna. Then the received signal can be shown as

$$\begin{aligned} \mathbf{Y} &= \mathbf{W}_{ABF}^H \mathbf{H} \mathbf{U}_{ABF} \mathbf{X} + \mathbf{W}_{ABF}^H \mathbf{Z} \\ &= \mathbf{W}_{ABF}^H \mathbf{A}_R \mathbf{D}_\alpha \mathbf{A}_T^H \mathbf{U}_{ABF} \mathbf{X} + \mathbf{W}_{ABF}^H \mathbf{Z}, \end{aligned} \quad (5)$$

where  $\mathbf{U}_{ABF}$  and  $\mathbf{W}_{ABF}$  are the analog precoding and combining matrices, respectively. They are obtained from the DFT codebook.  $\mathbf{Z}$  is Gaussian white noise with zero mean and unit variance. Due to the received signal passing through the ABF receiving matrix, the channel information suffers from a certain degree of loss. This dimension-deficient property makes channel estimation in HBF systems more challenging than traditional channel estimation.

### III. ALGORITHM DESIGN

Thanks to its wide-ranging applications in sparse signal processing, compressive sensing is widely used in sparse channel estimation [9], [16], [18], [19]. While CS algorithms may exhibit the floor effect when used in channel estimation [19], the floor effect appears when the channel length is close to or larger than the guard interval (GI). Ding *et al.* [19] used the prior known pseudo noise (PN) to replace the cyclic prefix (CP) and part of the PN is used to estimate the channel. Consequently, when the channel length (i.e. the maximum channel delay) is close to or larger than the guard interval (the length of the PN), there will be limited PN that can be used to estimate the channel. In other words, the PN during the channel delay time cannot be used for channel estimation, since the PN has been contaminated by the inter-block interferences. So if the channel length is close to or larger than the GI, there will be no more PN left to estimate the channel.

This scenario arises in digital television/terrestrial multimedia broadcasting (DTMB), which is not common in wireless communication. So, we will only discuss scenarios where the channel length is smaller than the GI.

As can be seen from Fig. 3, mmWave channels are naturally sparse, which means CS technology is a nature fit for estimating mmWave channels. The overcomplete dictionary has a major place in CS algorithms. The overcomplete dictionary of contrast algorithm in Section III-A is constructed using a universal method, while our proposed algorithm in Section III-B constructs the dictionary in an adaptive manner. Compared with the contrast algorithm, the proposed adaptive algorithm is slightly more complex, but the system performance is better, which is shown in Section IV.

#### A. CONTRAST ALGORITHM

CS channel estimation needs to first construct an overcomplete dictionary of the sparse channel. The conventional overcomplete dictionary  $\Theta = \Phi \mathbf{V}$  is composed of a random measurement matrix  $\Phi$ , such as a Gaussian matrix or a random  $\pm 1$  matrix, and a basis matrix  $\mathbf{V}$  [12], [15]. This approach is of low complexity and fast computation.

Based on the approach above, we propose a simple CS algorithm for HBF channel estimation as a contrast. As the arrival and departure steering matrices are unknown, we use virtual channel representation [16], [17]. It follows from the property of Kronecker product that the desired receive signal in (5) can be expressed as

$$\begin{aligned} \mathbf{Y}_{desired} &= \mathbf{W}_{ABF}^H \mathbf{A}_R \mathbf{D}_\alpha \mathbf{A}_T^H \mathbf{U}_{ABF} \mathbf{X} \\ &= \left( \mathbf{W}_{ABF}^H \mathbf{A}_R \right) \otimes \left( \mathbf{A}_T^H \mathbf{U}_{ABF} \mathbf{X} \right)^T \text{vec} \left( \mathbf{D}_\alpha \right) \\ &= \left( \mathbf{W}_{ABF}^H \mathbf{V}_R \right) \otimes \left( \mathbf{V}_T^H \mathbf{U}_{ABF} \mathbf{X} \right)^T \mathbf{h}_S, \end{aligned} \quad (6)$$

where  $\mathbf{h}_S$  is the vectorization of  $\mathbf{D}_\alpha$  and a sparse vector containing the elements of the virtual channel matrix  $\mathbf{H}_V$ . Then we vectorize the received signal  $\mathbf{Y}$  as

$$\begin{aligned} \text{vec}(\mathbf{Y}) &= \mathbf{G} \mathbf{h}_S + \text{vec} \left( \mathbf{W}_{ABF}^H \mathbf{Z} \right) \\ \mathbf{G} &= \left( \mathbf{W}_{ABF}^H \mathbf{V}_R \right) \otimes \left( \mathbf{V}_T^H \mathbf{U}_{ABF} \mathbf{X} \right)^T. \end{aligned} \quad (7)$$

Because matrix  $\mathbf{G} \in \mathbb{C}^{N_{RF} M_r \times M_r X_L}$  does not satisfy the Restricted Isometry Property (RIP), so  $\mathbf{G}$  cannot be used as the dictionary for estimating the channel directly. Baraniuk [12] show that an  $M \times N$  i.i.d. Gaussian matrix can be shown to have the RIP with high probability, if  $M \geq cK \log(N/K)$  with  $c$  being a small constant, where  $K$  is the sparsity of the signal to be estimated. If the Gaussian matrix is multiplied by an orthonormal basis, the product also satisfies the RIP with high probability. So letting  $M = N_{RF} N_T$  and  $N = N_R L$ , decomposing  $\mathbf{G}$  using SVD, and ignoring the noise, (7) can be rewritten as

$$\text{vec}(\mathbf{Y}) = \mathbf{U}_1 \mathbf{D} \mathbf{V}_1^H \mathbf{h}_S, \quad (8)$$

where  $\mathbf{U}_1 \in \mathbb{C}^M$ ,  $\mathbf{D} \in \mathbb{C}^{M \times N}$  and  $\mathbf{V}_1 \in \mathbb{C}^N$ . Premultiplying  $\mathbf{U}_1^H$  and  $\mathbf{D}^\dagger$  on both sides of (8) removes  $\mathbf{U}_1 \mathbf{D}$  on the right side.



$\mathbf{D}^\dagger$  is the pseudo-inverse matrix of  $\mathbf{D}$ .  $\mathbf{V}_1^H$  is an orthonormal basis, so we premultiply a Gaussian matrix  $\Phi_1$  on the both sides of (8) to obtain

$$\Phi_1 \mathbf{D}^\dagger \mathbf{U}_1^H \text{vec}(\mathbf{Y}) = \Phi_1 \mathbf{V}_1^H \mathbf{h}_S. \quad (9)$$

Denoting the observation signal by  $\tilde{\mathbf{y}} = \Phi_1 \mathbf{D}^\dagger \mathbf{U}_1^H \text{vec}(\mathbf{Y})$  and the overcomplete dictionary by  $\Theta_1 = \Phi_1 \mathbf{V}_1^H$ , we have

$$\tilde{\mathbf{y}}_1 = \Theta_1 \mathbf{h}_S. \quad (10)$$

The reconstruction of  $\mathbf{h}_S$  from  $\tilde{\mathbf{y}}$  can be done via " $l_0$ -norm" minimization, which can be expressed by

$$\begin{aligned} \hat{\mathbf{h}}_S &= \arg \min_{\mathbf{h}_S} \|\mathbf{h}_S\|_{l_0} \\ \text{s.t. } \tilde{\mathbf{y}}_1 &= \Theta_1 \mathbf{h}_S. \end{aligned} \quad (11)$$

### B. ADAPTIVE OVERCOMPLETE DICTIONARY CS

In contrast with the CS algorithm above, we propose an adaptive overcomplete dictionary for dimension-deficient channel estimation. In order to satisfy the RIP, there is a new way to build the overcomplete dictionary. According to [14], we have the fixed basis matrix  $\mathbf{V}_2 = \mathbf{V}_1^H$ . It should be noted that one can preselect a basis matrix  $\mathbf{V}_2$  different from this paper. Then matrix  $\Phi_2$  can be designed to satisfy  $\Theta_2^H \Theta_2 \approx \mathbf{I}$ , where  $\Theta_2 = \Phi_2 \mathbf{V}_2$ . We assume that

$$\Theta_2^H \Theta_2 = \mathbf{V}_2^H \Phi_2^H \Phi_2 \mathbf{V}_2 \approx \mathbf{I}. \quad (12)$$

Multiplying  $\mathbf{V}_2^H$  and  $\mathbf{V}_2$  on both sides of (12), we have

$$\mathbf{V}_2 \mathbf{V}_2^H \Phi_2^H \Phi_2 \mathbf{V}_2 \mathbf{V}_2^H \approx \mathbf{V}_2 \mathbf{V}_2^H. \quad (13)$$

If  $\mathbf{V}_2 \mathbf{V}_2^H$  has the eigenvalue decomposition  $\mathbf{V}_2 \mathbf{V}_2^H = \mathbf{U}_2 \Lambda \mathbf{U}_2^H$ , then (13) becomes

$$\mathbf{U}_2 \Lambda \mathbf{U}_2^H \Phi_2^H \Phi_2 \mathbf{U}_2 \Lambda \mathbf{U}_2^H \approx \mathbf{U}_2 \Lambda \mathbf{U}_2^H. \quad (14)$$

Then letting  $\Gamma = \Phi_2 \mathbf{U}_2$ , where  $\Gamma = [\boldsymbol{\tau}_1, \dots, \boldsymbol{\tau}_M]^T$  and  $\boldsymbol{\tau} = [\tau_1, \dots, \tau_N]^T$ , the problem becomes equivalent to optimizing  $\hat{\Gamma}$

$$\hat{\Gamma} = \arg \min \left\| \Lambda - \Lambda \Gamma^H \Gamma \Lambda \right\|_F^2. \quad (15)$$

Then  $\Phi_2$  can be generated from  $\hat{\Phi}_2 = \Gamma \hat{\mathbf{U}}_2^H$ . Consequently, we could obtain the overcomplete dictionary by using  $\hat{\Theta}_2 = \hat{\Phi}_2 \mathbf{V}_2$ .

Let  $[\lambda_1, \dots, \lambda_N]$  be the eigenvalues of the diagonal matrix  $\Lambda$ , which are ordered in decreasing order, such that  $\lambda_1 > \lambda_2 > \dots > \lambda_N$ . Then  $\mathbf{v}_i$  can be expressed as the elements of diagonal matrix  $\Lambda$  multiplied by the  $i$ th row of  $\Gamma$ , i.e.,

$$\mathbf{v}_i = [\lambda_1 \tau_{i,1}, \dots, \lambda_N \tau_{i,N}]^T. \quad (16)$$

Defining  $\mathbf{E}_j = \Lambda - \sum_{i=1, i \neq j}^m \mathbf{v}_i \mathbf{v}_i^T$ , then (15) can be rewritten as

$$\left\| \Lambda - \sum_{i=1, i \neq j}^m \mathbf{v}_i \mathbf{v}_i^T - \mathbf{v}_j \mathbf{v}_j^T \right\|_F^2 = \left\| \mathbf{E}_j - \mathbf{v}_j \mathbf{v}_j^T \right\|_F^2. \quad (17)$$

Let matrix  $\mathbf{E}_j$  be eigen-decomposed, i.e.,  $\mathbf{E}_j = \mathbf{B}_j \boldsymbol{\Sigma}_j \mathbf{B}_j^T$ , where  $\boldsymbol{\Sigma} = \text{diag}(\xi_1, \dots, \xi_M)$ , and  $\mathbf{B} = [\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_M]^T$ . Obviously, the objective of the optimization is to minimize the right hand side of (17). So let  $\mathbf{v}_j = \sqrt{\xi_{\max,j}} \boldsymbol{\beta}_{\max,j}$ , where  $\xi_{\max,j}$  is the maximum eigenvalue of  $\mathbf{E}_j$  and  $\boldsymbol{\beta}_{\max,j}$  is the corresponding eigenvector, and thus the primary gap of  $\mathbf{E}_j - \mathbf{v}_j \mathbf{v}_j^T$  can be eliminated. Then  $\boldsymbol{\tau}_j = [\tau_{j,1}, \dots, \tau_{j,N}]^T$  is updated as follows

$$[\lambda_1 \hat{\tau}_{j,1}, \dots, \lambda_N \hat{\tau}_{j,N}]^T = \sqrt{\xi_{\max,j}} \boldsymbol{\beta}_{\max,j}. \quad (18)$$

Then we can obtain the optimal  $\hat{\Gamma}$  using  $\boldsymbol{\tau}_j = \hat{\tau}_j$ . Furthermore, matrix  $\hat{\Phi}_2$  and dictionary  $\Theta_2$  are expressed as

$$\begin{aligned} \hat{\Phi}_2 &= \Gamma \hat{\mathbf{U}}_2^H \\ \hat{\Theta}_2 &= \hat{\Phi}_2 \mathbf{V}_2. \end{aligned} \quad (19)$$

To construct the newly observed signal according to (19), the original received signal in (8) can be rewritten as

$$\begin{aligned} \hat{\Gamma} \mathbf{U}_2^H \mathbf{D}^\dagger \mathbf{U}_1^H \text{vec}(\mathbf{Y}) &= \hat{\Gamma} \mathbf{U}_2^H \mathbf{V}_2 \mathbf{h}_S \\ \tilde{\mathbf{y}}_2 &= \hat{\Theta}_2 \mathbf{h}_S, \end{aligned} \quad (20)$$

where  $\tilde{\mathbf{y}}_2 = \hat{\Gamma} \mathbf{U}_2^H \mathbf{D}^\dagger \mathbf{U}_1^H \text{vec}(\mathbf{Y})$  and  $\hat{\Theta}_2 = \hat{\Gamma} \mathbf{U}_2^H \mathbf{V}_2$ , which is calculated in **Algorithm 1, Part I**. Through employing the compressive sampling match pursuit (CoSaMP) algorithm for CS reconstruction via (20), we obtain the sparse vector  $\mathbf{h}_S$ . Thus, the estimated channel can be written as

$$\hat{\mathbf{H}} = \mathbf{V}_R \hat{\mathbf{H}}_V \mathbf{V}_T^H, \quad (21)$$

where  $\hat{\mathbf{H}}_V$  is the de-vectorization of  $\mathbf{h}_S$  (in **Algorithm 1**,  $\mathbf{h}_S$  is replaced by  $\hat{\mathbf{h}}_{iK}$ ).

The notation in **Part II** in **Algorithm 1** is explained as follows.  $\mathbf{r}_i$  is the residual and  $i$  is the number of iterations.  $\Omega_i$  is the set of indexes (i.e., column numbers).  $\Theta_i$  is the set of columns from  $\hat{\Theta}$ , and the columns are selected by  $\Omega_i$ .  $\mathbf{J}_0$  means the set of  $2K$  indexes selected from each iteration. This is different from orthogonal matching pursuit (OMP). In other words, the traditional OMP algorithm chooses only one maximum value and the corresponding index from each iteration. By contrast, CoSaMP chooses  $2K$  maximum values and the corresponding indexes from each iteration. This is an improvement over OMP, which helps reduce the influence of accidental errors such as noise.

To summarize, the proposed algorithm consists of the following two steps, which are detailed in Algorithm 1:

- (1) Design the overcomplete dictionary  $\hat{\Theta}$ ;
- (2) Estimate the channel using CoSaMP.

Assuming the number of rows of  $\hat{\Theta}$  is far less than that of columns, i.e.,  $M \ll N$ , the complexity of Algorithm 1 is  $O(MN^3) \approx O(N^3)$ . This complexity is similar to that in [11].

### IV. SIMULATION RESULTS

In this section, the proposed adaptive algorithm is compared with the ones in [8], [10], and [18]. The UE is equipped with  $N_T = 32$ , the gNB is configured with  $N_R = 16$  antennas, and

**Algorithm 1** Overcomplete Dictionary Design & CoSaMP-Based Channel Estimation

**Require:**  $\mathbf{G}, K, M, \tilde{\mathbf{y}}, \mathbf{A}_T, \mathbf{A}_R$

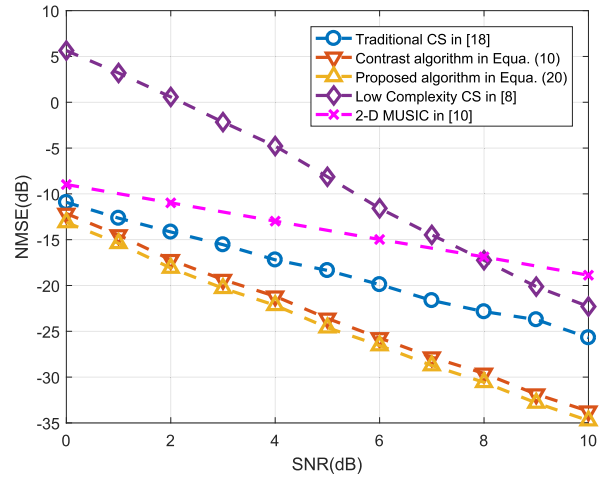
**Ensure:**  $\hat{\mathbf{H}}$

- 1: % **Part I** — Overcomplete dictionary design
  - 2: Basis matrix  $\mathbf{V}_2$ , a random initial matrix  $\Phi$ .
  - 3: Calculate the eigen decomposition  $\mathbf{V}_2\mathbf{V}_2^H = \mathbf{U}_2\mathbf{\Lambda}\mathbf{U}_2^H$ ;  $\Gamma = \Phi\mathbf{U}_2$ .
  - 4: **for** each  $i \in [1, M]$  **do**
  - 5: Calculate the eigenvalues of  $\mathbf{E}_j$
  - 6: Find the maximum eigenvalue  $\xi_{\max,j}$  and the corresponding eigenvector  $\beta_{\max,j}$
  - 7: Update the components of  $\Theta_j$  using (18)
  - 8: **end for**
  - 9: Calculate the optimal  $\Phi = \hat{\Phi}\mathbf{U}_2^H, \hat{\Theta} = \Phi\hat{\mathbf{V}}_2$ .
  - 10: % **Part II** — CE-CoSaMP
  - 11:  $\mathbf{r}_0 = \tilde{\mathbf{y}}, \Omega_0 = \emptyset, \Theta_0 = \emptyset$ .
  - 12: **for** each  $i \in [1, K]$  **do**
  - 13: Calculate  $\mathbf{u} = \hat{\Theta}^H \mathbf{r}_{i-1} = [\hat{\theta}_1, \dots, \hat{\theta}_N]^H \mathbf{r}_{i-1}$
  - 14: Find the maximum  $2K$  values of  $\mathbf{u}$ , and let the corresponding column indexes  $j_{2K}$  constitute set  $\mathbf{J}_0$
  - 15: Let  $\Omega_i = \Omega_{i-1} \cup \mathbf{J}_0, \Theta_i = \Theta_{i-1} \cup \hat{\theta}_j$  (forall  $j \in \mathbf{J}_0$ )
  - 16: Calculate  $\hat{\mathbf{h}}_i = \arg \min_{\mathbf{h}_i} \|\tilde{\mathbf{y}} - \Theta_i \mathbf{h}_i\| = (\Theta_i^T \Theta_i)^{-1} \Theta_i^T \tilde{\mathbf{y}}$
  - 17: Find the maximum  $K$  values of  $\hat{\mathbf{h}}_i$  and let them be  $\hat{\mathbf{h}}_{iK}$ ; the corresponding  $K$  columns of  $\Theta_i$  be the  $\Theta_{iK}$ ; the columns indexes of  $\hat{\Theta}$  be  $\Omega_{iK}$
  - 18: Update  $\Omega_i = \Omega_{iK}, \mathbf{r}_i = \mathbf{Q} - \Theta_{iK} \hat{\mathbf{h}}_{iK} = \mathbf{Q} - \Theta_{iK} (\Theta_i^T \Theta_i)^{-1} \Theta_i^T \tilde{\mathbf{y}}$
  - 19: If  $\mathbf{r}_i < 10^{-15}$  (which means the residual is small enough).
  - 20: **end for**
- De-vectorize
- 21: the  $\hat{\mathbf{h}}_{iK}$  into  $\hat{\mathbf{H}}_V$
  - 22: Calculate  $\hat{\mathbf{H}} = \mathbf{A}_R \hat{\mathbf{H}}_V \mathbf{A}_T^H$

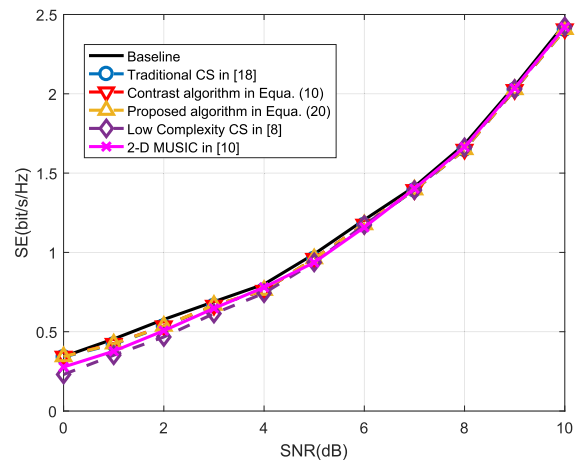
the number of chains  $N_{RF} = 2, \dots, 6$ . The virtual channel sparsity is  $K = L = 6$ , which means the non-zero value of  $\mathbf{H}_V$  is 6. We chose the virtual channel without leakage in our simulation, because the sparsity of the virtual channel with leakage is much larger than the number of channel paths ( $K > L$ ) and usually is not clear. The NMSE is adopted as the performance metric, which is defined as

$$\text{NMSE} = 10 \log_{10} \frac{\|\hat{\mathbf{H}} - \mathbf{H}\|_F^2}{\|\mathbf{H}\|_F^2}. \quad (22)$$

Fig. 5 depicts the NMSEs of five comparative channel estimation schemes. The number of RF chains in this simulation case is fixed to 6. The circle, rhombus, star, inverted triangle and triangle lines indicate the performances of the CE algorithm of traditional CS in [18], the 2-D MUSIC algorithm in [10], the low-complexity algorithm in [8], the contrast algorithm based upon (10) and the proposed algorithm based



**FIGURE 5.** NMSEs of the comparative algorithms.



**FIGURE 6.** SEs of the comparative algorithms versus the baseline with perfect CSI.

upon (20), respectively. As can be observed from the figure, the algorithm proposed in this paper outstrips the others in the sense of the NMSE. The algorithm in [8] has the worst performance and is almost 15 dB worse than its comparative counterparts at low SNRs. This is because the algorithm in [8] estimates the sparse channel coefficient matrix and the two angle matrices (AOA and AOD) at the same time. The estimation errors caused by noise have greater influence than the other algorithms, especially at low SNRs. The 2-D MUSIC and traditional CS algorithms perform very close to the proposed two algorithms in this paper at low SNRs, but the gap grows when SNR increases. While 2-D MUSIC is slightly worse than the traditional CS method, because the method to estimate the path gain used in [10] is least-square (LS). In other words, the LS algorithm is sensitive to noise and the estimation results are easily affected by the noise. Furthermore, the results demonstrate that the CS algorithm with the adaptive dictionary sees a certain improvement in NMSE over the universal dictionary scheme.

Fig. 6 plots the SE of the five comparative methods and the baseline with perfect CSI. The results indicate that the five methods perform nearly identically at high SNRs, while the method in [8] has a certain gap at low SNRs. This is because

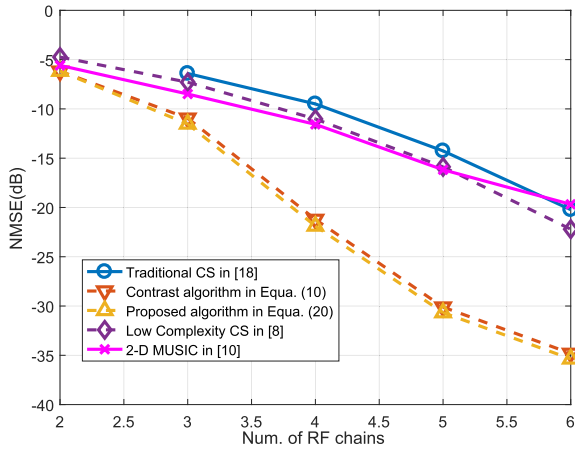


FIGURE 7. NMSEs of the comparative algorithms versus the number of RF chains.

the channel estimation results in [8] are not as good as those of the other methods, which is shown in Fig. 5.

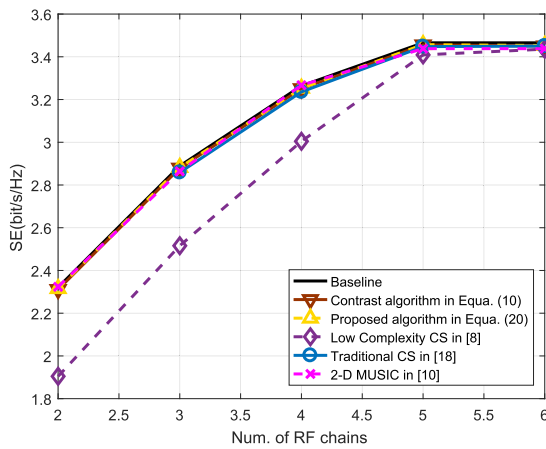


FIGURE 8. SEs of the comparative algorithms versus the number of RF chains.

Fig. 7 and 8 show the effects of the number of RF chains on the NMSE and the SE at SNR = 10 dB. We do not consider the traditional CS algorithm in [18] when  $N_{RF} = 2$ , because there are errors in calculating  $\hat{h}_i$  when  $N_{RF} = 2$ . When the number of RF chains is two, the rank of the ABF matrix equals two, i.e.,  $\text{rank}(\mathbf{U}_{ABF}^T) = \text{rank}(\mathbf{W}_{ABF}^H) = 2$ . Then the rank of the sensing matrix  $\text{rank}(\hat{\Phi}) = \text{rank}(\mathbf{U}_{ABF}^T \otimes \mathbf{W}_{ABF}^H) = 4 < K$ . Consequently, matrix  $\Theta_i$  (in Algorithm 1, line 15) is singular matrix when updating it and the calculation of  $\hat{h}_i$  will be wrong because matrix  $\Theta_i^T \Theta_i$  is not full rank and cannot be inverted. Fig. 7 indicates that the proposed algorithm reaps nearly 30 dB gains in NMSE, when the number of RF chains grows from 2 to 6. While the SE in Fig. 8 converges to the upper band after the number of RF chains becomes more than 5. This is because the data streams limit the growth of SE in the network.

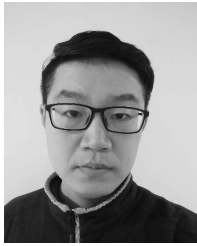
## V. CONCLUSION

In this paper, we proposed a defective-dimension channel estimation algorithm based on CS technology. Compared with non-adaptive algorithms, the proposed algorithm

designed an overcomplete dictionary adaptively with the objective of improving the channel estimation performance in terms of NMSE. The proposed scheme also shows a better noise robustness performance than the existing methods, and the SE is close to the upper bound of perfect CSI. Our future research will focus on two more challenging scenarios. One is CS channel estimation for virtual channels with adjacent leakage, while the other one is when sparsity is not *a priori* known.

## REFERENCES

- [1] J. Nightingale, J. M. A. Calero, Q. Wang, and P. Salva-Garcia, "5G-QoE: QoE modelling for ultra-HD video streaming in 5G networks," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 621–634, Jun. 2018.
- [2] 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.
- [3] T. S. Rappaport et al., "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.
- [4] F. Sbrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.
- [5] 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.
- [6] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.
- [7] 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.
- [8] H.-L. Chiang, T. Kadur, G. Fettweis, and W. Rave, "Low-complexity spatial channel estimation and hybrid beamforming for millimeter wave links," in *Proc. IEEE Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Valencia, Spain, Sep. 2016, pp. 942–948.
- [9] Y. Han and J. Lee, "Asymmetric channel estimation for multi-user millimeter wave communications," in *Proc. IEEE Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Jeju, South Korea, Oct. 2016, pp. 4–6.
- [10] Z. Guo, X. Wang, and W. Heng, "Millimeter-wave channel estimation based on 2-D beamspace MUSIC method," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5384–5394, Aug. 2017.
- [11] M. L. Bencheikh, Y. Wang, and H. He, "Polynomial root finding technique for joint DOA DOD estimation in bistatic MIMO radar," *Signal Process.*, vol. 90, no. 9, pp. 2723–2730, Mar. 2010.
- [12] R. G. Baraniuk, "Compressive sensing [lecture notes]," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–124, Jul. 2007.
- [13] T. E. Bogale, L. B. Le, and X. Wang, "Hybrid analog-digital channel estimation and beamforming: Training-throughput tradeoff," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 5235–5249, Dec. 2015.
- [14] J. M. Duarte-Carvajalino and G. Sapiro, "Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1395–1408, Jul. 2009.
- [15] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [16] J. Sung, J. Choi, and B. L. Evans, (2017). "Wideband channel estimation for hybrid beamforming millimeter wave communication systems with low-resolution ADCs." [Online]. Available: <https://arxiv.org/abs/1710.10669>
- [17] A. M. Sayeed, "Deconstructing multi-antenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.
- [18] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.
- [19] W. Ding, F. Yang, C. Pan, L. Dai, and J. Song, "Compressive sensing based channel estimation for OFDM systems under long delay channels," *IEEE Trans. Broadcast.*, vol. 60, no. 2, pp. 313–321, Jun. 2014.
- [20] A. F. Molisch et al., "Hybrid beamforming for massive MIMO: A survey," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, Sep. 2017.
- [21] I. Ahmed et al., "A survey on hybrid beamforming techniques in 5G: Architecture and system model perspectives," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3060–3097, 4th Quart., 2018.



**YU XIAO** received the B.S. degree from Xidian University, Xi'an, China, in 2011, and the M.S. degree from the North China University of Technology, Beijing, China, in 2016. He is currently pursuing the Ph.D. degree with the Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, China. His research interests include hybrid beamforming, channel estimation, and compressive sensing.



**YAFENG WANG** (S'00–M'03–SM'09) received the B.Sc. degree from the Baoji University of Arts and Science, in 1997, the M.Eng. degree from the University of Electronic Science and Technology of China, in 2000, and the Ph.D. degree from the Beijing University of Posts and Telecommunications, in 2003. In 2008, he was a Visiting Scholar with the Faculty of Engineering and Surveying, University of Southern Queensland, Australia.

He is currently a Professor of electronic engineering with the School of Information and Telecommunications, Beijing University of Posts and Telecommunications. He leads the Broadband Mobile Communication Engineering Laboratory, which is one of the Zhongguancun Science Park Open Laboratories. He has authored or co-authored over 100 peer-reviewed journal and conference papers. His research mainly focuses on wireless communications and information theory.



**WEI XIANG** (S'00–M'04–SM'10) received the B.Eng. and M.Eng. degrees in electronic engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 1997 and 2000, respectively, and the Ph.D. degree in telecommunications engineering from the University of South Australia, Adelaide, Australia, in 2004. From 2004 to 2015, he was with the School of Mechanical and Electrical Engineering, University of Southern Queensland, Toowoomba, Australia. He is currently the Founding Professor and the Head of the Discipline of the Internet of Things Engineering with the College of Science and Engineering, James Cook University, Cairns, Australia. He has authored or co-authored over 200 peer-reviewed journal and conference papers. His research interests include the broad areas of communications and information theory, particularly the Internet of Things, and coding and signal processing for multimedia communications systems. He is an Elected Fellow of the IET and the Engineers Australia. He was named as a Queensland International Fellow by the Queensland Government of Australia, from 2010 to 2011, as an Endeavour Research Fellow by the Commonwealth Government of Australia, from 2012 to 2013, as a Smart Futures Fellow by the Queensland Government of Australia, from 2012 to 2015, and as a JSPS Invitational Fellow jointly by the Australian Academy of Science and the Japanese Society for Promotion of Science, from 2014 to 2015. He was a Finalist of the 2016 Pearcey Queensland Award. He was a recipient of the TNQ Innovation Award, in 2016. He was a co-recipient of three best paper awards at 2015 WCSP, 2011 IEEE WCNC, and 2009 ICWMC. He received several prestigious fellowship titles. He is the Vice Chair of the IEEE Northern Australia Section. He was an Editor of *IEEE COMMUNICATIONS LETTERS*, from 2015 to 2017. He is an Associate Editor of *Telecommunications Systems* (Springer). He has served for a large number of international conferences as the General Co-Chair, as the TPC Co-Chair, and as the Symposium Chair.

• • •