

RESEARCH

Open Access



DFT codebook-based hybrid precoding for multiuser mmWave massive MIMO systems

Yu Huang^{1,2,3}, Chen Liu^{1,2*}, Yunchao Song^{1,2} and Xiaolei Yu³

Abstract

In millimeter wave (mmWave) massive MIMO (multiple-input multiple-output) systems, it is difficult to apply conventional digital precoding techniques due to hardware constraints. Fortunately, the hybrid precoding can be utilized to reduce power consumption and high costs. In this paper, a codebook-based hybrid precoding scheme for downlink multiuser mmWave massive MIMO systems is proposed. Our main idea is that the analog and digital precoders are designed separately to maximize the achievable sum rate. In the analog domain, we take the potential multiuser conflict and angular domain of channel matrix into consideration and propose an efficient conflicting-aware (CA) beam-column selection method to obtain a discrete Fourier transform (DFT) codebook-based analog precoder. According to the CA method, all users are classified into two groups, i.e., conflicting users (CUs) and non-conflicting users (NCUs). Different criteria of beam-column selection are applied for the two user groups. Then, zero-forcing (ZF) digital precoder is directly used in the digital domain. Simulation results illustrate that our proposed algorithm which has low complexity achieves satisfactory SR performance, which approaches that of the full digital precoding (the upper bound) and outperforms other existing hybrid algorithms.

Keywords: mmWave, Massive MIMO, Hybrid precoding, DFT, Codebook

1 Introduction

Millimeter-wave (mmWave) massive MIMO has been considered as a promising candidate for the fifth generation of mobile telecommunications (5G) [1], since it will enable gigabits per second transmitting data rate owing to the large bandwidth available in mmWave spectrum. In order to implement high-quality communication in mmWave massive MIMO systems, the deployment of a large number of antenna arrays are required at base stations (BSs). Each BS needs to serve multiple mobile stations (MSs) simultaneously for efficient system performance. The precoding technology that is applied to generate transmitting signals at BSs is usually processed in the baseband for conventional lower frequency

systems. Nevertheless, for conventional digital precoding, energy-intensive radio frequency (RF) chains are required for each antenna. The energy consumption of each RF chain (about 250 mW per RF chain [2]) is a large portion of the total energy consumption. If traditional full digital precoding is used in mmWave massive MIMO systems, the relevant large number of RF chains will produce high energy consumption. The high hardware cost of numerous RF chains makes traditional fully digital precoders unfit for mmWave massive MIMO systems unfortunately [3]. So precoding in the above systems is a trend to be divided into analog domain and digital domain due to current research.

The authors in [3] investigated hybrid precoding which contains analog and digital precoding for single-user scenario. It was indicated that hybrid precoding technology can reduce the number of required RF chains and performs almost the same as fully digital precoding [3] was designed to obtain spatial multiplexing gain, and it can achieve good performance only when numerous antennas are deployed at both transmitter and receiver. The

* Correspondence: liuch@njupt.edu.cn

¹College of Electronic and Optical Engineering & College of Microelectronics, Nanjing University of Posts and Telecommunications, Nanjing, China

²National and Local Joint Engineering Laboratory of RF Integration and Micro-Assembly Technology, Nanjing University of Posts and Telecommunications, Nanjing, China

Full list of author information is available at the end of the article

beam steering scheme for the multiuser scenario was proposed in [4], which firstly chose the best beam steering vectors as the analog precoding matrix for each user and then applied the zero-forcing (ZF) precoding in the digital domain. This algorithm performed well in single-path scenarios, but would deteriorate a lot in multi-path scenarios because beam steering could not make full use of multipath channel gains.

Another related set of works has considered the codebook-based hybrid precoding, where a precoder is selected from a set of a priori fixed codebook. The complexity of the codebook-based design is generally much lower than the non-codebook-based design. The codebooks in the IEEE 802.15.3c standard [5] of indoor systems and IEEE 802.11ad [6] standard of wireless local area networks (WLAN) were designed. A Q-bit resolution codebook using the optimal precoding weight vector was proposed in [7] to achieve a uniform maximum gain in all directions. However, the column vectors in codebooks of [5–7] are not orthogonal, which may lead to performance degradation for multiuser scenarios due to the multiuser interference. For mmWave systems, an efficient discrete Fourier transform (DFT) codebook-based precoding training scheme was proposed in [8]. The analog precoding matrix is produced by column vector selection from the DFT codebook whose columns are orthogonal [9]. proposed hybrid precoding design, where mutually unbiased bases combined (MUB) were applied for digital precoder and the DFT codebook was invoked for analog precoder. A single-user scenario was considered in [9], which had a good performance and low complexity. For multiuser scenario, the DFT codebook-based beam selection scheme was proposed in [10] to formulate the hybrid precoding design into an optimization problem, referred to exhausted searching which traverses all the possible beam combinations. The joint optimization of the DFT codebook-based hybrid precoding was investigated in [11] for multiuser mmWave systems under two performance measures: sum rate (SR) and energy efficiency (EE). The joint codeword selection and precoder design (JWSPD) scheme in [11] are flexible, which can achieve a satisfactory performance for both SR and EE.

In this paper, we propose a new codebook-based hybrid precoding to maximize the achievable SR for multiuser mmWave massive MIMO systems. Especially, the analog and digital precoders are designed separately. In the analog domain, the constant modulus limitation of analog precoder should be taken into account. So the analog precoder is derived from the DFT codebook whose column vectors are orthogonal and each entry of which is constant modulus. We propose an efficient conflicting-aware (CA) beam selection scheme for column selection from the DFT codebook of analog

precoder, with the consideration of the potential multiuser conflict. Note that the same beam column in DFT codebook may be selected for different users, which will cause severe multiuser interference. Moreover, some RF chains are likely to be wasted because they do not contribute to SR. Hence, all users are firstly classified into two groups, i.e., non-conflicting users (NCUs) and the conflicting users (CUs). If the dominant beam index of one user differs from any other user's, the user is defined as NCU. On the contrary, the user whose dominant beam index is the same as any other users is defined as CU. Then, different criteria of beam-column selection are applied for the two user groups. For NCUs, the beams with the largest signal to leakage plus noise ratio (SLNR) are picked out, while for CUs, the best beams are chosen by an incremental algorithm with low complexity in order to maximize the achievable SR. The proposed CA searching scheme performs similarly to the exhausted searching algorithm in [10]. In the digital domain, zero-forcing (ZF) technique [12] is adopted to obtain the digital precoder. An arbitrary number of RF chains of the selection is considered in this paper without the restriction that the number of RF chains equals the number of users. Finally, the numerical results validate that our proposed codebook-based hybrid precoding scheme achieves satisfactory SR performance, which approaches to that of the full digital precoding (the upper bound), and gets a better performance than other existing hybrid precoding algorithms, such as [4, 5, 7, 11]. The contributions of the paper are summarized as follows.

- The hybrid precoding design is investigated in multiuser mmWave massive MIMO systems under SR performance optimization with non-convex constraint. Our scheme is based on the idea that analog and digital precoders are separately designed.
- Considering the constant modulus limitation of analog precoder, we propose a DFT codebook-based analog precoder, whose orthogonal beam columns (i.e., analog precoding vectors) are specified by DFT codeword.
- We propose a novel CA beam-column selection method for designing analog precoder, which can efficiently classify users into two groups, i.e., CUs and NCUs. Different criteria of beam-column selection are applied for the two user groups.
- Our proposed codebook-based hybrid precoding scheme has low computational complexity. The channel matrix is mapped from the spatial domain to angular domain. Since the channel matrix in the spatial domain is sparse, the mapping can decrease the complexity while selecting beams. Then, due to the Sherman-Morrison-Woodbury identity [13], a

recursive way can be used on the matrix inversion which can decrease the complexity.

The remaining of this paper is organized as follows. Section 2 introduces the system model and formulates the SR maximization problem. The proposed DFT codebook-based hybrid precoding design with CA beam selection method and complexity analysis are shown in Section 3. In Section 4, simulations are provided to evaluate the performance of the proposed scheme. Conclusions are presented in Section 5.

1.1 Notation

In this paper, boldface letters denote vectors and matrices; $E(\cdot)$ denotes the expectation; $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^{-1}$ denote the transpose, conjugate transpose, inversion, respectively; $|\cdot|$ denotes the determinant of a matrix; $\|\cdot\|_F$ denotes the Frobenius norm of a matrix; $\text{diag}(p)$ denotes the diagonal matrix. $\mathcal{E}^{M \times M}$ represents the spaces of $M \times M$ matrices with complex entries; $\mathcal{CN}(m, R)$ denotes the complex Gaussian distribution with mean m and covariance R ; I_U is the $U \times U$ identity matrix. $\mathcal{A} \setminus \mathcal{B}$ denotes the set where the elements in set \mathcal{B} are eliminated from set \mathcal{A} . Finally, $[\cdot]$ represents to round a decimal to its nearest lower integer.

2 System model and problem statement

2.1 System and signal model

In this paper, a fully connected downlink mmWave system is considered, which is shown in Fig. 1. Every RF chain is connected to all antennas of the BS. Now, a general scenario with one BS serving U users simultaneously is considered. The BS is equipped with N_{RF} RF chains

and N_{BS} antennas. Each MS is equipped with N_{MS} antennas. The BS communicates with each MS only through one stream. So the total number of transmitting streams is $N_S = U$. In addition, the number of MS cannot exceed the number of BS RF chains [7], i.e., $N_S \leq N_{\text{RF}}$. The spatial multiplexing gain of this hybrid system for multiuser is limited by $N_{\text{BS}} \geq N_{\text{RF}}$.

On the downlink multiuser mmWave MIMO system as shown in Fig. 1, the BS applies an $N_{\text{RF}} \times N_S$ digital precoding matrix $\mathbf{F}_{\text{BB}} = [\mathbf{f}_{\text{BB}}^1, \mathbf{f}_{\text{BB}}^2, \dots, \mathbf{f}_{\text{BB}}^U]$ where $\mathbf{f}_{\text{BB}}^u (u \in [1, 2, \dots, U])$ is an $N_{\text{RF}} \times 1$ vector, followed by an $N_{\text{BS}} \times N_{\text{RF}}$ RF precoding matrix. The transmitting signal is therefore

$$\mathbf{x} = \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}} \mathbf{s}, \quad (1)$$

where $\mathbf{s} = [s_1, s_2, \dots, s_U]^T$ is the message to be transmitted which is an $N_S \times 1 (U \times 1)$ vector and normalized as $E[\mathbf{s}\mathbf{s}^H] = (1/N_S)\mathbf{I}_U$. Since \mathbf{F}_{RF} is the analog precoding matrix implemented as analog phase shifters, and the entries of \mathbf{F}_{RF} are constant modulus, which should be normalized to satisfy $|F_{\text{RF}}[i, j]| = 1/\sqrt{N_{\text{BS}}}$, $\forall i = 1, 2, \dots, N_{\text{BS}}, j = 1, 2, \dots, N_{\text{RF}}$, where $F_{\text{RF}}[i, j]$ is the $(i, j)^{\text{th}}$ element of analog precoding matrix \mathbf{F}_{RF} . The total power constraint of the hybrid precoding matrix is enforced by normalizing \mathbf{F}_{BB} as $\|\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}\|_F^2 = N_S$.

The received signal observed by the u^{th} MS is

$$\mathbf{y}_u = \sqrt{\rho} \mathbf{H}_u \sum_{n=1}^U \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB}}^n s_n + \mathbf{n}_u, \quad (2)$$

where \mathbf{H}_u is an $N_{\text{MS}} \times N_{\text{BS}}$ matrix which indicates the channel between the BS and the u^{th} user, $\mathbf{n}_u \in \mathcal{CN}(0, \sigma^2)$

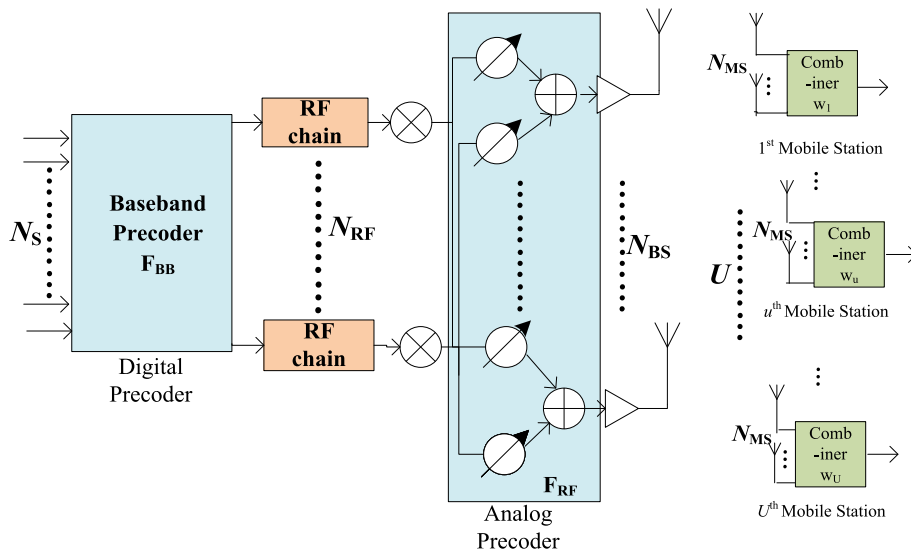


Fig. 1 A multiuser mmWave massive MIMO system model

\mathbf{I}_{N_r}) represents the Complex Gaussian noise vector, and ρ represents the average received power.

At the u^{th} MS, the received signal is processed by the u^{th} combiner \mathbf{w}_u . The processed signal is

$$\mathbf{y}_u = \mathbf{w}_u^H \sqrt{\rho} \mathbf{H}_u \sum_{n=1}^U \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB}}^n \mathbf{s}_n + \mathbf{w}_u^H \mathbf{n}_u, \quad (3)$$

where $\mathbf{w}_u \in \mathbb{C}^{N_{\text{MS}} \times 1}$ is the digital combining vector of the u^{th} user.

2.2 Millimeter-wave channel model

An extended Saleh-Valenzuela model is used in this paper due to the high antenna correlation and the limited spatial selectivity, which is commonly applied in mmWave massive MIMO channel modeling [14–16]. We consider the scenario that each scatter reflects one path. An adoption of a geometric channel model with L_u scatters is used for the channel of the u^{th} user. Under this model, the channel matrix of the u^{th} user $\mathbf{H}_u \in \mathbb{C}^{N_{\text{MS}} \times N_{\text{BS}}}$ can be expressed as

$$\mathbf{H}_u = \sqrt{\frac{N_{\text{MS}} N_{\text{BS}}}{L_u}} \sum_{l=1}^{L_u} \alpha_{u,l} \mathbf{a}_{u,\text{MS}}(\theta_{u,l}) \mathbf{a}_{u,\text{BS}}^H(\phi_{u,l}), \quad (4)$$

where N_{BS} is the number of antennas at the BS, N_{MS} is the number of antennas at the u^{th} MS. $\alpha_{u,l}$ is the complex gain of the l th path, which includes the path loss with $E[|\alpha_{u,l}|^2] = \bar{\alpha}$ and $\bar{\alpha}$ is a normalization constant. The variable $\phi_{u,l} \in [0, 2\pi]$ is the l th path's angle of departure (AoD). The variable $\theta_{u,l} \in [0, 2\pi]$ is the l th path's angle of arrival (AoA). At last, $\mathbf{a}_{u,\text{BS}}(\phi_{u,l})$ is the antenna array response vector of the BS. $\mathbf{a}_{u,\text{MS}}(\theta_{u,l})$ is the antenna array response vector of the u^{th} MS. L_u is the number of propagation paths for the channel of the u^{th} MS with $l \in [1, 2, \dots, L_u]$. We assume the BS and every MS get knowledge about the structure of their antenna arrays. A typical uniform linear array (ULA) is utilized in the simulations. If a ULA is equipped, $\mathbf{a}_{u,\text{BS}}(\phi_{u,l})$ is defined as

$$\mathbf{a}_{u,\text{BS}}(\phi_{u,l}) = \frac{1}{\sqrt{N_{\text{BS}}}} \left[1, e^{j \frac{2\pi d}{\lambda} \sin(\phi_{u,l})}, \dots, e^{j (N_{\text{BS}}-1) \frac{2\pi d}{\lambda} \sin(\phi_{u,l})} \right]^T, \quad (5)$$

where d is the distance between adjacent antenna elements and λ is the wavelength of mmWave. $\mathbf{a}_{u,\text{MS}}(\theta_{u,l})$ of the MS can be described in a similar pattern. If each MS has one antenna, no AoAs exist. Therefore, (4) can be simplified as

$$\mathbf{h}_u = \sqrt{N_{\text{BS}}/L_u} \sum_{l=1}^{L_u} \alpha_{u,l} \mathbf{a}_{u,\text{BS}}^H(\phi_{u,l}).$$

Now, \mathbf{H} indicates the collective channel matrix from all users, which is denoted as

$$\mathbf{H} = [\mathbf{H}_1^T, \mathbf{H}_2^T, \dots, \mathbf{H}_U^T]^T. \quad (6)$$

The received signal vector $\mathbf{y} = [\mathbf{y}_1, \mathbf{y}_1, \dots, \mathbf{y}_U]^T$ of the total U users can be expressed as

$$\mathbf{y} = \mathbf{w}^H \sqrt{\rho} \mathbf{H} \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}} \mathbf{s} + \mathbf{w}^H \mathbf{n}, \quad (7)$$

where $\mathbf{n} \in \mathcal{CN}(0, \sigma^2 \mathbf{I}_{U \times N_{\text{MS}}})$ represents the Complex Gaussian noise vector and $\mathbf{w} = [\mathbf{w}_1^T, \mathbf{w}_2^T, \dots, \mathbf{w}_U^T]^T$ is the collective combiner of all users.

The channel model can be translated into the beam space or angular domain [17] via a beamforming matrix \mathbf{U}_t , which is an $N_{\text{BS}} \times N_{\text{BS}}$ unitary matrix. The columns of \mathbf{U}_t are the vectors in the DFT matrix with uniform spacing. The (k, l) entry of \mathbf{U}_t is analytically defined as

$$\mathbf{U}_t(k, l) = \frac{1}{\sqrt{N_{\text{BS}}}} \exp\left(\frac{-j2\pi kl}{N_{\text{BS}}}\right), k, l = 0, \dots, N_{\text{BS}}-1. \quad (8)$$

The DFT codebook that divides the space into N_{BS} orthogonal beams is introduced. An N_{BS} -dimensional DFT matrix can be denoted as

$$\mathbf{U}_t = \frac{1}{\sqrt{N_{\text{BS}}}} \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & e^{j \frac{2\pi}{N_{\text{BS}}}} & \dots & e^{j \frac{2\pi (N_{\text{BS}}-1)}{N_{\text{BS}}}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{j \frac{2\pi (N_{\text{BS}}-1)}{N_{\text{BS}}}} & \dots & e^{j \frac{2\pi (N_{\text{BS}}-1)^2}{N_{\text{BS}}}} \end{bmatrix}. \quad (9)$$

Hence, \mathbf{H}^a is defined as an equivalent representation of the channel in angular domain [17], which can be denoted as

$$\mathbf{H}^a = \mathbf{H} \mathbf{U}_t. \quad (10)$$

(7) can be expressed in the angular domain as follows.

$$\mathbf{y} = \mathbf{w}^H \sqrt{\rho} \mathbf{H}^a \mathbf{U}_t^H \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}} \mathbf{s} + \mathbf{w}^H \mathbf{n}. \quad (11)$$

The relationships between the channel matrix of the angular domain \mathbf{H}^a and channel matrix of the spatial domain \mathbf{H} in (6) are denoted as

$$\mathbf{H}^a \mathbf{U}_t^H = \mathbf{H}, \quad (12)$$

$$\mathbf{H}^a = [\mathbf{U}_t^H \mathbf{H}_1^T, \mathbf{U}_t^H \mathbf{H}_2^T, \dots, \mathbf{U}_t^H \mathbf{H}_U^T]^T. \quad (13)$$

The channel matrix \mathbf{H}^a is a mapping of the signals for each MS in a new domain of orthogonal vectors. The channel matrix in the spatial domain \mathbf{H} is the discrete Fourier transform of the channel matrix in the angular domain \mathbf{H}^a . The \mathbf{U}_t defined in (8) is unitary, i.e., $\mathbf{U}_t \mathbf{U}_t^H = \mathbf{U}_t^H \mathbf{U}_t = \mathbf{I}$. Moreover, the representation of signals in the angular domain which can be applied to other antenna array structures is a more general concept. Since the number of dominant scatters is very limited in

mmWave systems, the non-zero elements of every vector in angular channel is much smaller than N_{BS} , namely, the angular domain channel matrix \mathbf{H}^a is sparse [18]. This sparse structure can be utilized to design \mathbf{F}_{RF} with dimension reduced by column vector selection from the DFT codebook \mathbf{U}_t whose columns are orthogonal. The ideal performance can only be achieved when the users located in the orthogonal subspace in mmWave massive MIMO systems. The multiplication of \mathbf{U}_t and \mathbf{F}_{RF} can obtain parallel sub-channels to transmit signals after analog precoding according to (11). Moreover, \mathbf{U}_t satisfies the constant modulus constraint which is enforced by the feasible domain of analog precoder \mathbf{F}_{RF} . Then, the problem of designing \mathbf{F}_{RF} is transformed to select proper columns from the DFT codebook.

2.3 Problem formulation

The objective is to efficiently design appropriate hybrid precoding matrices at the BS for the mmWave massive MIMO system. For this purpose, the common performance measures are considered: the achievable sum rate (SR) of the system. Our optimum target is to maximize SR in the downlink multiuser scenario. Assume that the BS can get the perfect channel state information (CSI). According to the processed signal received at the u^{th} user in (3), the achievable data rate of the u^{th} user is

$$R_u = \log_2 \left(1 + \frac{\frac{\rho}{N_S} |\mathbf{w}^H \mathbf{H}_u \mathbf{F}_{RF} \mathbf{f}_{BB}^u|^2}{\frac{\rho}{N_S} \sum_{n \neq u} |\mathbf{w}^H \mathbf{H}_u \mathbf{F}_{RF} \mathbf{f}_{BB}^n|^2 + \sigma^2} \right). \quad (14)$$

The SR of the system is

$$R_{\text{SUM}} = \sum_{u=1}^U R_u. \quad (15)$$

Although other criterion such as the max-min fairness criterion is interesting in [19], we aim to maximize the achievable SR in this paper, since the achievable SR is a crucial criterion in the mmWave systems. The SR performance is preferable to fairness when precoders of BS are practically operated, which brings greater system capacity and better user experience. Hence, we choose SR performance as the optimization objective.

In our paper, we assume that each MS is equipped with a small number of antennas, since N_{MS} is usually much smaller than N_{BS} in massive MIMO systems [16]. The BS communicates with each MS only through one stream. Therefore, in order to get the multi-antenna gain, digital combining is utilized at the MSs. The hybrid combining cannot reduce RF chains when the number of MS antennas is smaller than the number of

propagation paths in the channel. The digital combiner of MS is practical when each MS has a few antennas, since it will not bring plenty of RF chains and high cost of hardware overhead. The matched filter is chosen by us as the combiner at MSs, since the matched filter is an optimal linear filter, and its output signal to interference plus noise ratio (SINR) can be maximized which can make a great contribution to the system sum rate. The matched filtering (MF) combiner is defined as

$$\mathbf{w}_u = \hat{\mathbf{H}}_u, \quad \text{with MF}, \quad (16)$$

where $\hat{\mathbf{H}}_u$ is an effective channel matrix which can be indicated as $\hat{\mathbf{H}}_u = \mathbf{H}_u \sum_{n=1}^U \mathbf{F}_{RF} \mathbf{f}_{BB}^n$.

The proceeding design problem of \mathbf{F}_{RF} and \mathbf{F}_{BB} can be described as

$$\begin{aligned} \mathbf{F}_{RF}^{\text{OPT}}, \mathbf{F}_{BB}^{\text{OPT}} &= \underset{\mathbf{F}_{RF}, \mathbf{F}_{BB}}{\operatorname{argmax}} R_{\text{SUM}}, \\ \text{s.t. } \mathbf{F}_{RF} &\in \mathbf{W}, \\ \|\mathbf{F}_{RF} \mathbf{F}_{BB}\|_F^2 &= N_S \end{aligned} \quad (17)$$

where \mathbf{W} represents the feasible domain of analog precoders that is the set of $N_{BS} \times N_{RF}$ matrices with constant-magnitude entries ($|\mathbf{F}_{RF}[i, j]| = 1/\sqrt{N_{BS}}$). The total power constraint of the BS is enforced by normalizing \mathbf{F}_{BB} as $\|\mathbf{F}_{RF} \mathbf{F}_{BB}\|_F^2 = N_S$. No other hardware-related constraints are imposed on the digital precoding matrix.

The optimal solution to (17) is almost impossible to obtain due to several difficulties, including the RF constraints, the couple between \mathbf{F}_{RF} and \mathbf{F}_{BB} . However, near-optimal solutions can be found through iterative algorithms, which require a large number of training processes and feedback overhead. Therefore, a direct solution to this maximization problem of sum rate in (17) is neither tractable nor practical. In view of the application difficulties of previous precoding algorithms in mmWave, we develop a new DFT codebook-based hybrid algorithm suitable for multiuser mmWave massive MIMO systems in Section 3. Our proposed scheme achieves better performance than the other algorithms.

3 Methods

In this section, we propose a codebook-based hybrid precoding algorithm to solve the optimization problem and analyze its computational complexity. The main idea of our scheme is dividing the calculation of the hybrid precoding into analog and digital domains.

3.1 The proposed hybrid algorithm

In the digital domain, zero-forcing (ZF) technique is adopted to obtain the digital precoder, based on the equivalent channel matrix.

In the analog domain, the analog precoder \mathbf{F}_{RF} is picked from DFT codebook [9] which is also called beam selection. Beams mean column vectors of the DFT codebook. Our target is to select N_{RF} best unshared beams from total N_{BS} beams for U users to maximize the achievable sum rate. Hence, it is crucial to choose proper beams which will be applied for the design of analog precoder, although that can be done by exhaustive searching with high time cost over all the possible beam combinations to identify the optimum beams. A more efficient CA beam selection method that can make the balance between performance and time consumption is proposed for the column vector selection of the analog precoder. Then, ZF precoding is directly utilized for digital precoder. CA method consists of two stages: (1) group to identify NCUs and CUs by consideration of the potential multiuser conflict and (2) fix the beam indexes for NCUs and seek the best exclusive beam indexes for CUs, in order to maximize the SR. As \mathbf{F}_{RF} is the abstraction from the DFT matrix \mathbf{U}_D , the column indexes of DFT codebook which can be picked by \mathbf{F}_{RF} are obtained. For the sake of simplicity, the scenario that each MS has one antenna and $N_{\text{RF}} = U$ will be firstly explained as follows.

3.1.1 Stage 1: Grouping stage to identify CUs and NCUs

Our goal is to pick N_{RF} best dedicated beams from the total N_{BS} beams of \mathbf{U}_t to maximize the achievable sum rate. As we know, the optimal solution can be obtained by exhaustive search, which has the best performance but suffers the high complexity which is $C_{N_{\text{BS}}}^{N_{\text{RF}}}$. For a mmWave massive MIMO system with $N_{\text{BS}} = 128$ and $N_{\text{RF}} = 16$, the total searching number reaches $128! / 16! (128 - 16)! = 9.8 \times 10^{18}$. Hence, a more practical selection scheme to get the near-optimal performance is necessary.

Firstly, we sort the row elements of \mathbf{H}^a in ascending order of the ratio T as follows.

$$T_{u,i} = \frac{|\mathbf{H}^a(u,i)|^2}{\sum_{j=1, j \neq i}^{N_{\text{BS}}} |\mathbf{H}^a(u,j)|^2 + \sigma^2}, u = 1, 2, \dots, U, i = 1, 2, \dots, N_{\text{BS}}, \quad (18)$$

where $\mathbf{H}^a(u,i)$ is the $(u,i)^{\text{th}}$ element of the matrix \mathbf{H}^a . The maximum value $T_{u,i}$ in the whole u^{th} row of the angular channel \mathbf{H}^a means the dominant beam of the u^{th} user. The dominant beam index i of the u^{th} user is denoted as $b_u^* \in [1, 2, \dots, N_{\text{BS}}]$. The dominant beams $\{b_u^*\}_{u=1}^U$ make a direct contribution to achievable sum rate according to Appendix. Hence, if $b_1^* \neq b_2^* \neq \dots \neq b_U^*$, the beam set $\mathcal{B} = \{b_1^*, b_2^*, \dots, b_U^*\}$ can get the near optimal sum rate. But if there are two users sharing the same

dominant beam, the two users will suffer severe multi-user interference. Unfortunately, if U is large and N_{BS} is small, the probability of selecting the same dominant beam is obviously large. Therefore, users can be grouped. Users are categorized into line-of-sight (LOS) and non-line-of-sight (NLOS) groups in [20], based on the channel conditions. In this paper, all U users are classified into two groups, i.e., non-conflicting users (NCUs) and conflicting users (CUs) according to their dominant beam index in \mathbf{H}^a . Different criteria to select beams are adopted for different user groups. An example for grouping is given as follows.

- (1) User u is defined as NCU, if the dominant beam index of u differs from any other users, i.e., $b_u^* \notin \{b_1^*, \dots, b_{u-1}^*, b_{u+1}^*, \dots, b_U^*\}$. The NCU group \mathcal{F}_{NCU} consists of all NCU users. The example is shown in Fig. 2a, in which the dominant beam index of users can be obtained according to (18) as $b_1^* = 2, b_4^* = 8$ and $b_2^* = b_3^* = 5$. It is obvious that $\mathcal{F}_{\text{NCU}} = \{1, 4\}$, since $b_1^* \notin \{b_2^*, b_3^*, b_4^*\}$ and $b_4^* \notin \{b_1^*, b_2^*, b_3^*\}$. The dominant beam index can be directly selected for the user $u \in \mathcal{F}_{\text{NCU}}$, because this beam not only make a direct contribution to achievable sum rate but also produce little interference to other users.
- (2) User u is defined as CU, if the dominant beam index of u is the same as any other users, i.e., $b_u^* \in \{b_1^*, \dots, b_{u-1}^*, b_{u+1}^*, \dots, b_U^*\}$. The CU group \mathcal{F}_{CU} consists of all CU users. $\mathcal{F}_{\text{CU}} \cup \mathcal{F}_{\text{NCU}} = \{1, 2, \dots, U\}$ and $\mathcal{F}_{\text{CU}} \cap \mathcal{F}_{\text{NCU}} = \emptyset$ are obvious. The example in Fig. 2a is considered, where $\mathcal{F}_{\text{CU}} = \{2, 3\}$, since $b_2^* \in \{b_1^*, b_3^*, b_4^*\}$ and $b_3^* \in \{b_1^*, b_2^*, b_4^*\}$. For the user $u \in \mathcal{F}_{\text{CU}}$, the suitable beam index is picked from the set $\{1, 2, \dots, N_{\text{BS}}\} \setminus \{b_u^* | u \in \mathcal{F}_{\text{NCU}}\}$, which is introduced in the next stage as shown in Fig. 2b.

A simple example for single-antenna MS is taken above to better understand our proposed scheme. If each MS of the user has several antennas, the angular channel of the u^{th} user is not a vector, but a matrix denoted as \mathbf{H}_u^a . Then, we search all elements in \mathbf{H}_u^a , instead of the row elements of \mathbf{H}^a , and the ratio T is sorted in ascending order as well.

3.1.2 Stage 2: Best beam indexes searching stage

For the user $u (u \in \mathcal{F}_{\text{NCU}})$, the best beam index can be directly selected as the dominant beam index. In the above example, the best indexes of NCUs are $\text{Card}(\mathcal{F}_{\text{NCU}}) = \{2, 8\}$. Then, another search for $u \in \mathcal{F}_{\text{CU}}$ remains $(\text{Card}(N_{\text{BS}}) - \text{Card}(\mathcal{F}_{\text{NCU}})) = \{1, 3, 4, 5, 6, 7\}$ beams to maximize the SR. A classical ZF precoding technology is adopted in the digital domain as follows.

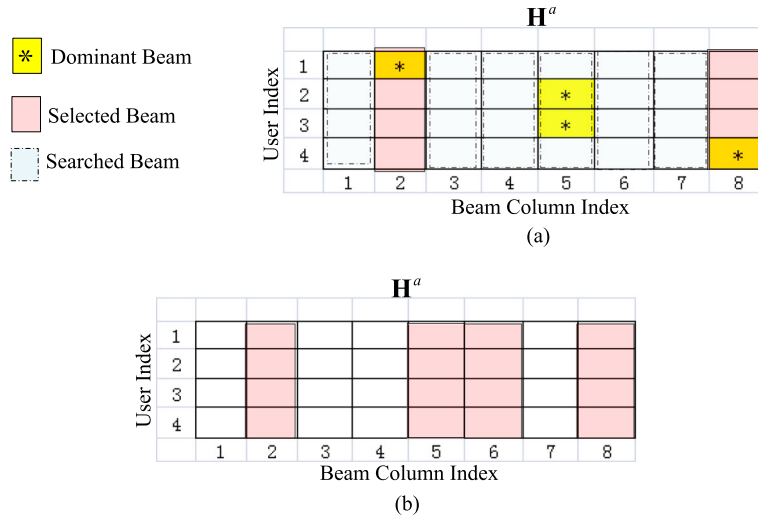


Fig. 2 An example of the CA beam-column selection method. **a** Stage 1: Grouping stage to identify CUs and NCUs. **b** Stage 2: Best beam searching stage

$$\mathbf{F}_{\text{BB}} = \alpha \mathbf{H}_{\text{eq}}^H \left(\mathbf{H}_{\text{eq}} \mathbf{H}_{\text{eq}}^H \right)^{-1}, \quad (19)$$

where $\mathbf{H}_{\text{eq}} = \mathbf{H}^a \mathbf{U}_t^H \mathbf{F}_{\text{RF}}$ is the equivalent channel matrix seen from the perspective of the digital domain. α is a scaling factor to make sure that $E\{\|\mathbf{F}_{\text{BB}}\mathbf{s}\|_2^2\} = \rho$. Hence, the factor α is depicted as

$$\alpha = \sqrt{\frac{\rho}{\text{tr}\left(\left(\mathbf{H}_{\text{eq}} \mathbf{H}_{\text{eq}}^H\right)^{-1}\right)}}. \quad (20)$$

If equal power allocation is utilized at the BS, the average data rate of the u^{th} user is given by

$$R_u = \log_2 \left(1 + \frac{|\alpha|^2}{N_s \sigma^2} \right). \quad (21)$$

Then, the SR maximization problem in (17) can be written as

$$\begin{aligned} \mathbf{F}_{\text{RF}}^{\text{OPT}} &= \underset{D}{\text{argmax}} R_{\text{SUM}} \\ &= \underset{D}{\text{argmax}} U \log_2 \left(1 + \frac{\rho}{N_s \sigma^2 \text{tr}\left(\mathbf{H}_{\text{eq}}^r \left(\mathbf{H}_{\text{eq}}^r\right)^H\right)^{-1}} \right), \end{aligned} \quad (22)$$

where D is the possible beam candidate set for the CUs with $\text{Card}(\mathcal{F}_{\text{NCU}})$, which is the unshared columns selected from the set $\{1, 2, \dots, N_{\text{BS}}\} \setminus \{b_u^* | u \in \mathcal{F}_{\text{NCU}}\}$. We denote $\mathbf{H}_{\text{eq}}^r = \mathbf{H}_{\text{eq}}(:, s), s \in D \cup \{b_u^* | u \in \mathcal{F}_{\text{NCU}}\}$.

It is obvious that maximizing the SR is equivalent to minimize $\text{tr}(\mathbf{H}_{\text{eq}}^r (\mathbf{H}_{\text{eq}}^r)^H)^{-1}$. Hence, (22) can be further formulated as

$$\mathbf{F}_{\text{RF}}^{\text{OPT}} = \underset{D}{\text{argmin}} \text{tr} \left(\mathbf{H}_{\text{eq}}^r \left(\mathbf{H}_{\text{eq}}^r \right)^H \right)^{-1}. \quad (23)$$

Next, the beam columns with the greatest contribution to the SR are picked out in each step. The optimal solution is by an exhaustive search which involves prohibitively high complexity of search with $C_{N_{\text{BS}}}^{N_{\text{RF}}}$ possible combinations. A low-complexity incremental algorithm [21] for beam-column selection is proposed to get a near-optimal solution.

As the example is shown as Fig. 2, in which CUs consist of user 2 and user 3, since $b_2^* = b_3^* = 5$. Therefore, in the first step, the unshared beam column b_2 of user 2 should be selected as

$$b_2 = \underset{b}{\text{argmin}} \text{tr}(\mathbf{G} + \mathbf{g}_b \mathbf{g}_b^H)^{-1}, \mathbf{G} \triangleq \mathbf{A} \mathbf{A}^H + \varepsilon \mathbf{I}, \quad (24)$$

where $\mathbf{A} = \mathbf{H}_{\text{eq}}(:, s), s \in \{b_u^* | u \in \mathcal{F}_{\text{NCU}}\}$ is a matrix of beam columns selected for NCUs, $\mathbf{g}_b = \mathbf{H}_{\text{eq}}(:, b), b \in S = \{1, 2, \dots, N_{\text{BS}}\} \setminus \{b_u^* | u \in \mathcal{F}_{\text{NCU}}\}$ is one beam column possibly selected for CUs. ε is set to make \mathbf{G} nonsingular, which can guarantee the matrix inversion in (24) exists. ε is a positive number (e.g., 10^{-2}).

Based on the Sherman-Morrison-Woodbury identity [13], a recursive way can be used to obtain the matrix inversion which can decrease the complexity. The matrix inversion can be obtained with complexity $\mathcal{O}(U^2)$ instead of $\mathcal{O}(U^3)$. A convenient expression is given for the

inverse of the matrix $(\mathbf{B} + \mathbf{U}\mathbf{V}^H)$ where \mathbf{B} is n by n and \mathbf{U} and \mathbf{V} are n by k :

$$(\mathbf{B} + \mathbf{U}\mathbf{V}^H)^{-1} = \mathbf{B}^{-1} - \mathbf{B}^{-1}\mathbf{U}(\mathbf{I} + \mathbf{V}^H\mathbf{B}^{-1}\mathbf{U})^{-1}\mathbf{V}^H\mathbf{B}^{-1}. \quad (25)$$

In (25), both \mathbf{B} and $(\mathbf{I} + \mathbf{V}^H\mathbf{B}^{-1}\mathbf{U})$ are assumed to be nonsingular. A rank- k correction to a matrix results in a rank- k correction of the inverse. Particularly, if $k = 1$, \mathbf{B} is nonsingular, $\beta = 1 + \mathbf{v}^H\mathbf{B}^{-1}\mathbf{u} \neq 0$, then

$$(\mathbf{B} + \mathbf{u}\mathbf{v}^H)^{-1} = \mathbf{B}^{-1} - \frac{1}{\beta}\mathbf{B}^{-1}\mathbf{u}\mathbf{v}^H\mathbf{B}^{-1}. \quad (26)$$

According to (26), since \mathbf{G} is nonsingular, (24) can be equivalently formulated as

$$b_2 = \underset{b}{\operatorname{argmin}} \operatorname{tr} \left(\mathbf{G}^{-1} - \frac{\mathbf{G}^{-1}\mathbf{g}_b\mathbf{g}_b^H\mathbf{G}^{-1}}{1 + \mathbf{g}_b^H\mathbf{G}^{-1}\mathbf{g}_b} \right). \quad (27)$$

By calculating the objective function in (27), b_2 has been selected. In the example in Fig. 2, user 2 selects the beam 5. Then, \mathbf{G} and S can be updated as $\mathbf{G} = \mathbf{G} + \mathbf{g}_{b_2}\mathbf{g}_{b_2}^H$ and $S = S \setminus b_2$. Then, user 3 in CUs uses the same method presented above to pick other beams. We assume that user 3 selects the beam 6 after computation. If there are more users, an incremental algorithm should continue until all $\operatorname{Card}(\mathcal{F}_{\text{CU}})$ beams have been picked out. Consequently, the set \mathcal{B}^* of all selected beam columns can be formed as

$$\mathcal{B}^* = \{b_1, \dots, b_{\operatorname{Card}(\mathcal{F}_{\text{CU}})}\} \cup \{b_u^* | u \in \mathcal{F}_{\text{NCU}}\} \quad (28)$$

In this paper, the number of RF chains selected is arbitrary. Many methods such as [4] solve the hybrid precoding problem, but $N_{\text{RF}} = U$ is strictly demanded, which cannot meet the general scene as $N_{\text{RF}} \geq U$. Our proposed algorithm satisfies the normal condition without the constraint $N_{\text{RF}} = U$. Hence, this can utilize the additional RF chains to explore all potential gains of sum rate. As mentioned above, we first introduce per-user CA beam selection method above that each of the user is assigned with one RF chain. While the condition is $N_{\text{RF}} \geq U$, each user will be assigned multiple RF chains. For fairness, we strive to make every user use the same number of RF chains. Denote $C = \lceil N_{\text{RF}}/U \rceil$, each user is firstly allocated with C RF chains which can be acquired from the CA method executing C rounds. The remaining RF chains can be denoted as $T = N_{\text{RF}} - \lceil N_{\text{RF}}/U \rceil \cdot U$. The remaining RF chains are prioritized to allocate to NCUs, since users without multiuser interference can bring higher SR. Then, two cases are classified according to the number of NCUs which is denoted as $N_{\mathcal{F}_{\text{NCU}}}$: (1) If $N_{\mathcal{F}_{\text{NCU}}} > T$, each of the first T th NCUs is allocated with one RF chains. (2) If $N_{\mathcal{F}_{\text{NCU}}} < T$, all of the NCUs are allocated

with one RF chain. Each of the first $(T - N_{\mathcal{F}_{\text{NCU}}})^{\text{th}}$ CUs is allocated with one RF chain.

3.2 Computational complexity

Here, we compare the computational complexities of our proposed algorithm with the full digital algorithm, the exhausted searching algorithm, and the algorithms in [4, 5, 7, 11], which are presented in Table 1. The complexity of the proposed algorithm includes two parts as follows.

- (1) The first part comes from the beam selection of the analog precoding matrix. If there are K CUs and $U - K$ NCUs, it requires $(N_{\text{BS}} - U) + (K + 1 - k)$, $k = 1, 2, \dots, K$, possible beam selections for the k th CU. The inversion of the matrix has a complexity of $\mathcal{O}(U^2)$.
- (2) The second part is from the calculation of ZF digital precoder. The complex multiplication of the matrix requires $\mathcal{O}(N_{\text{BS}}U^2)$.

Consequently, the totally computational complexity of the proposed algorithm is $\mathcal{O}(N_{\text{BS}}U^2 + U^2)$. Meanwhile, the complexity of the full digital algorithm is $\mathcal{O}(N_{\text{BS}}^3)$. Compared to the exhausted searching, the proposed CA scheme reduces the computation time from $C_{N_{\text{BS}}}^{N_{\text{RF}}}$ to U^2 , which is much more acceptable. Hence, the complexity of an exhausted searching hybrid algorithm is $\mathcal{O}(N_{\text{BS}}U^2 + C_{N_{\text{BS}}}^{N_{\text{RF}}})$. The complexity of beam steering hybrid algorithm in [4] is $\mathcal{O}(N_{\text{BS}}^2U^2)$. The complexities of hybrid algorithms in [5, 7] are $\mathcal{O}(N_{\text{BS}}U^2 + U^2)$ as well, since beam selection methods except for selected codebooks in them are the same in simulations of the next section. The complexity of the JWSPD hybrid algorithm in [11] is $\mathcal{O}(N_{\text{BS}}^4U^4)$.

It can be noticed that our proposed hybrid precoding algorithm can highly reduce the computational complexity compared with the full digital algorithm and the exhausted searching algorithm. Besides this above, when N_{BS} is much larger than U , the complexity of the

Table 1 Comparison of the computational complexities

Algorithm	Median operation
Proposed DFT hybrid	$\mathcal{O}(N_{\text{BS}}U^2 + U^2)$
Full digital algorithm	$\mathcal{O}(N_{\text{BS}}^3)$
Exhausted searching hybrid	$\mathcal{O}(N_{\text{BS}}U^2 + C_{N_{\text{BS}}}^{N_{\text{RF}}})$
Beamsteering hybrid in [4]	$\mathcal{O}(N_{\text{BS}}^2U^2)$
802.15.3c hybrid in [5]	$\mathcal{O}(N_{\text{BS}}U^2 + U^2)$
Q-bit resolution hybrid in [7]	$\mathcal{O}(N_{\text{BS}}U^2 + U^2)$
JWSPD hybrid in [11]	$\mathcal{O}(N_{\text{BS}}^4U^4)$

proposed algorithm is much lower than the algorithms in [4, 11]. Though the complexity of our proposed algorithm is equal to the algorithms in [5, 7], our proposed hybrid precoding can get better performance.

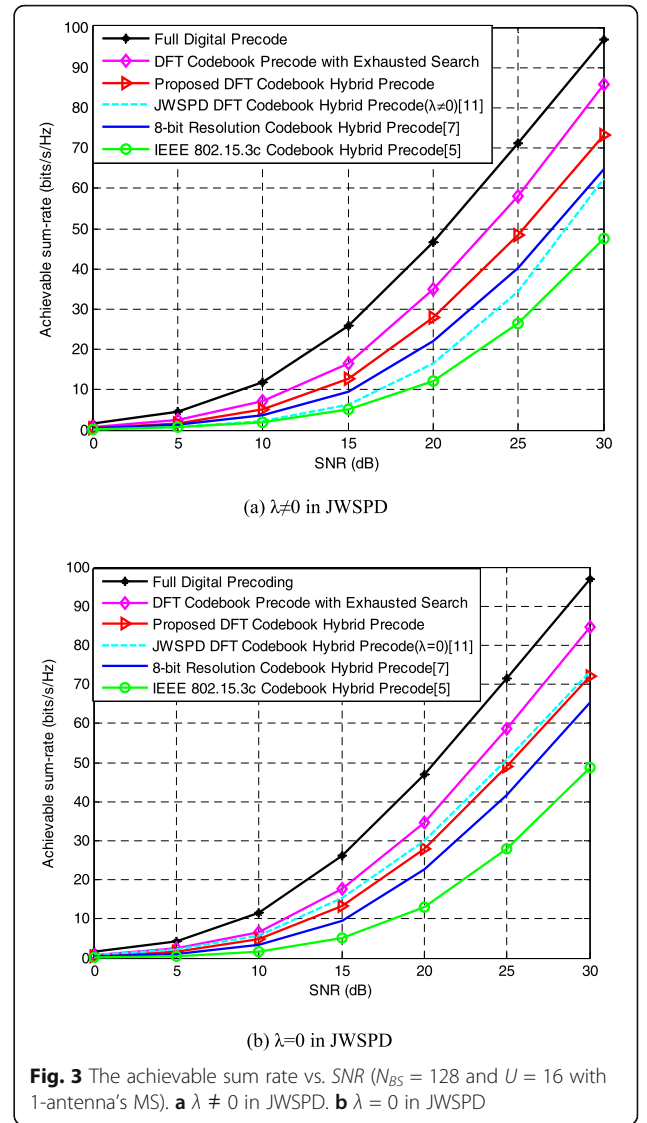
4 Results and discussion

In this section, we evaluate the performance of our proposed algorithm with the full digital precoding, the algorithm based on DFT codebook with exhausted searching scheme mentioned in [9], beam steering hybrid algorithm in [4], IEEE 802.15.3c codebook-based hybrid precoding [5], Q-bit resolution codebook-based hybrid precoding [7], and JWSPD hybrid algorithm in [11], which are all under fully connected architecture for fairness.

The simulation parameters are described as follows. Considering the system model depicted in Section 2, the ULA is implemented at the BS. The spacing between antenna elements of BS is half-wavelength. The channel of every MS has 5 propagation paths. The AoD of every propagation path is uniformly distributed $[0, 2\pi]$. The average power of every propagation path is randomly generated from a uniform random variable within $[0, 1]$. The total received power is normalized for every MS in order to satisfy the power constraint. Signal to noise ratio (SNR) is defined as $SNR = \rho/\sigma^2$. The simulation results are generated by 1000 trial Monte-Carlo simulation. The performance of precoding algorithms is shown in terms of achievable SR and EE in the following subsections.

4.1 Achievable sum rate

Figure 3 depicts the performance comparison of achievable sum rate, against the average SNR, with the BS antenna number $N_{BS} = 128$ and user number $U = 16$. A simple scenario is assumed that the number of RF chains is equal to the number of MSs and each MS is equipped with one antenna. It is shown that the performance of our proposed algorithm is close to the full digital algorithm (the upper bound), but the number of RF chains is significantly reduced. The proposed algorithm can also get better performance than the algorithms in [5, 7]. The comparison of our proposed hybrid precoding and the JWSPD scheme proposed in [11] is given under two different configurations, i.e., $\lambda \neq 0$ and $\lambda = 0$. A turnable sparse parameter λ is introduced in [11] to control the sparsity of the optimization solution. The larger λ is, the more sparse solution is. Thus, λ is properly chosen firstly in the JWSPD scheme to balance maximizing the system SR and minimizing the number of the selected virtual antennas (codewords). Then, the remained (unselected) RF chains with the corresponding phase shifter networks can be turned off to save power. In other words, the introduced sparse parameter λ represents a trade-off



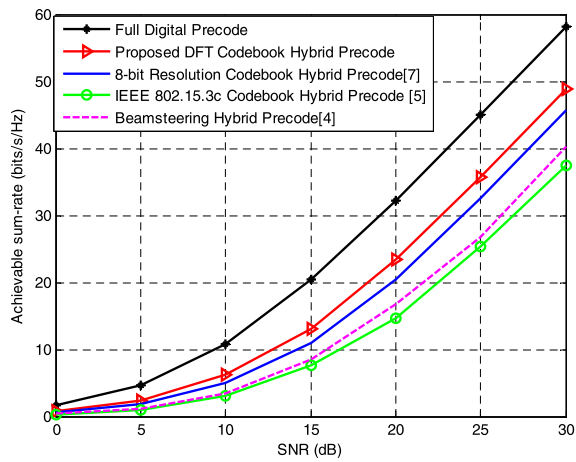
between system SR and EE. However, our proposed algorithm directly optimizes only one factor which is SR. Hence, the SR performance of our proposed scheme is better than the SRmax JWSPD scheme, when λ is properly chosen and $\lambda \neq 0$ as it is shown in Fig. 3a. If the sparse parameter λ is mandatorily chosen to zero in the simulation of Fig. 3b, the JWSPD scheme achieves a little better SR performance than our proposed scheme, but with higher computational complexity. Since $\lambda = 0$ means that the sparsity of the optimization solution and EE are mandatorily unconsidered, the system SR can be maximized.

The performance of our proposed algorithm is also close to the exhausted searching DFT codebook-based hybrid algorithm. The exhausted searching algorithm traverses all the possible beam combinations and jointly exploits the optimum beams. However, it is extremely

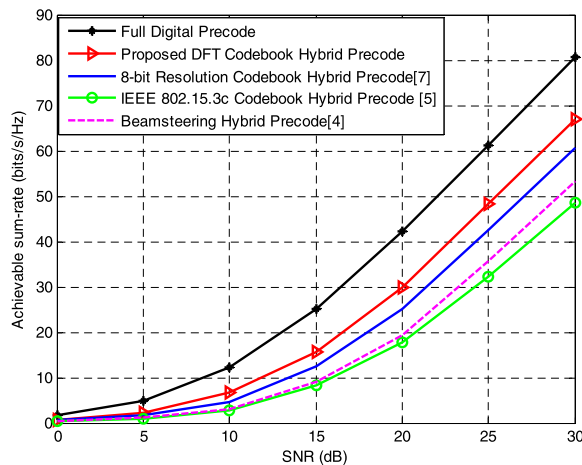
time-consuming when N_{BS} grows large. Our proposed scheme is a more practical scheme with an acceptable performance difference. When fixing the SR at 40 bits/s/Hz, the proposed hybrid algorithm can get 5 dB gain compared to [5]. That demonstrates the CA beam selection method and DFT codebook of the proposed algorithm are efficient and attractive ways to solve the optimal problem (17).

Figure 4 depicts the performance comparison of achievable sum rate, against the average SNR, with $U = 4$, and $N_{BS} = 128$ or $N_{BS} = 256$. Each MS is equipped with two antennas in simulations. The BS serves U MSs simultaneously using N_{RF} , and RF chains are assumed for simplicity as well.

As shown in Fig. 4a, the performance of our proposed algorithm is better than the algorithms in [4, 5, 7], when



(a) $N_{BS}=128$



(b) $N_{BS}=256$

Fig. 4 The achievable sum rate vs. SNR ($U = 4$ with 2-antennas' MS).

a $N_{BS} = 128$. **b** $N_{BS} = 256$

$N_{BS} = 128$. The performance difference is caused by the following three reasons. Firstly, the column vectors are not orthogonal in beam steering codebook in [4], 802.15.3c codebook in [5] and Q-bit resolution codebook in [7], which may lead to performance degradation due to multiuser interference. However, we use the DFT codebook which has orthogonal columns. Secondly, our proposed CA method for columns selection from the codebook is more flexible than the selection method in [4]. CA method utilizes the sparse angular domain of the channel matrix, with the consideration of the potential multiuser conflict. But the selection method in [4] is based on the assumption of a single path in the channel matrix. Thirdly, the algorithm in [4] performed well in single-path scenarios but would deteriorate a lot in multi-path scenarios because beam steering could not make full use of multi-path channel gains. In fact, mmWave systems are mainly based on multi-path scenarios in reality. Our proposed algorithm can perform well in both single-path and multi-path scenarios.

Some observations can be obtained by analyzing Fig. 4b. The sum rates of all the algorithms grow as the number of BS antennas N_{BS} increases from 128 to 256. The sum rate of our proposed algorithm almost rises to 70 bps/Hz at SNR is 30 dB. But the performance gaps between the proposed algorithm and other algorithms are almost invariable with the increasing number of N_{BS} , since numerous antennas are deployed at BS which can obtain spatial multiplexing gain.

Figure 5 depicts the performance comparison of achievable sum rate, against the average SNR, with different number of users at $N_{BS} = 256$. The number of user U is the same as the number of RF chains N_{RF} , for three different configurations such as $U = 32$, $U = 16$, and $U = 8$. From Fig. 5, it can be observed that the SR of the proposed algorithm improves with the increasing number of users. But the rate gap between the proposed

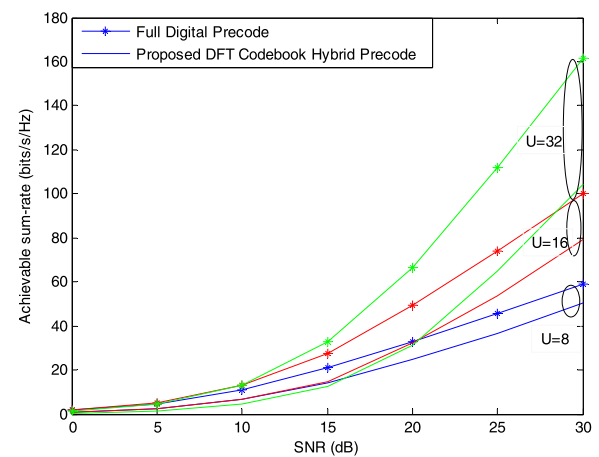


Fig. 5 The achievable sum rate vs. SNR with the number of users

algorithm and full digital algorithm becomes bigger as the number of users increases. This trend is because the inter-user interference becomes high.

Figure 6 depicts the performance comparison of achievable sum rate, against the number of RF chains at $N_{BS} = 128$, $U = 16$, and $SNR = 25$ dB. The number of RF chains is from 8 to 128. It can be observed that the SR of the proposed algorithm enhances by increasing RF chains owing to the diversity gains supplied by multiple RF chains. It can take full advantage of additional RF chains. Some other observations can be obtained by analyzing Fig. 6. The first is that there are some differences between full digital algorithm and the proposed algorithm at low N_{RF} . The second is the performance finally converges to the full digital algorithm (the upper bound) by increasing N_{RF} , while the gap reduces to zero at $N_{RF} = 128$. The third is when $N_{RF} = 64$, and the proposed algorithm achieves 99% achievable sum rate of the full digital algorithm.

Figure 7 depicts the performance comparison of the achievable sum rate, against the average SNR, with different power allocation schemes for users, i.e., waterfilling power distribution [22] and equal power allocation. For waterfilling, power allocation is implemented after hybrid precoding. The effective channel matrix $\hat{\mathbf{H}}_u$ is firstly implemented by singular value decomposition (SVD) to get parallel subchannels. Then, the Lagrangian method is used to distribute the power of subchannels. It is shown that waterfilling power distribution scheme is better than equal power allocation. The performance of our proposed hybrid precoding with waterfilling power distribution is much closer to the full digital precoding (the upper bound). Due to Lagrangian of waterfilling power distribution method can be used to distribute the power of subchannels, the system capacity is the largest. Consequently, the achievable sum rate of the system can be maximized.

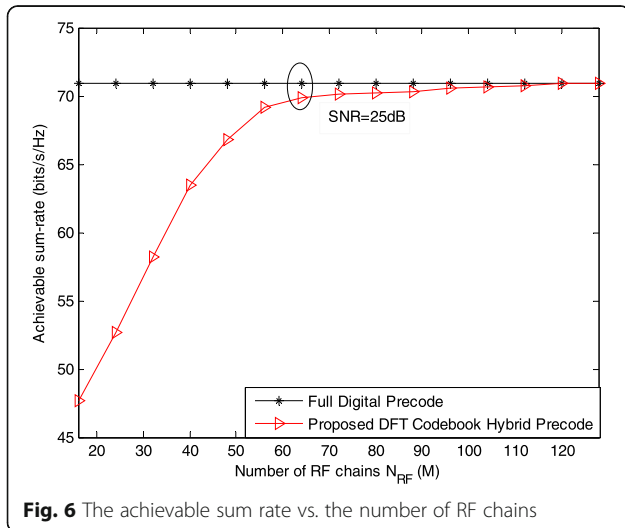


Fig. 6 The achievable sum rate vs. the number of RF chains

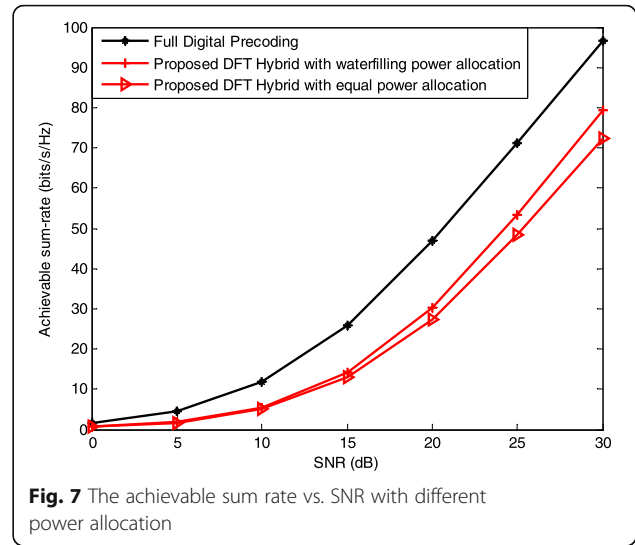


Fig. 7 The achievable sum rate vs. SNR with different power allocation

4.2 Energy efficiency

The energy efficiency ζ is defined in [2] as the ratio between the achievable sum rate and the total power consumption, i.e., $\zeta = \frac{R_{SUM}}{P + N_{RF}P_{RF} + P_{BB}}$ (bps/Hz/W), where P is the transmitting power of BS, P_{RF} is the power consumption of every RF chain, and P_{BB} is the power consumption of digital precoder. In general, the typical values are adopted that $P_{RF} = 250$ mW, and $P_{BB} = 300$ mW [23].

Figure 8 shows the performance of EE, against the average SNR, also with the BS antenna number $N_{BS} = 128$ and user number $U = 16$. It is shown that our proposed algorithm can get higher EE than other five compared algorithms. In particular, our proposed algorithm performs much better than the full digital precoding algorithm because N_{RF} equals to N_{BS} in full digital precoding. A large number of RF chains cause very high energy consumption, i.e., 250 mW per RF chain. In

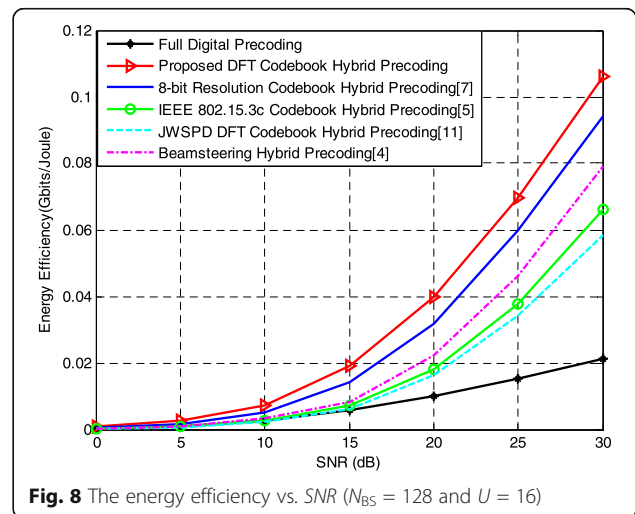


Fig. 8 The energy efficiency vs. SNR ($N_{BS} = 128$ and $U = 16$)

hybrid precoding, RF chains are far less than antennas at the transmitter. Therefore, the hybrid precoding algorithm can extremely reduce the energy consumption generated by the RF chains compared to the full digital scheme. Moreover, in terms of EE, our proposed algorithm outperforms the JWSPD scheme [11] ($\lambda = 0$), which achieves a similar SR performance to that of our scheme in Fig. 3b. Since EE are unconsidered in the JWSPD scheme as $\lambda = 0$.

Figure 9 shows the performance of EE, against the number of users, where SNR is 15 dB and the number of users increases from 5 to 25. It is obvious that EE of our proposed algorithm decreases with the increasing number of users, since more users will cause more RF chains selected at BS with more energy consumption generated. The EE of our proposed algorithm is higher than all of other five algorithms, especially when there are not too many users U .

5 Conclusion

In this paper, a low-complexity hybrid precoding scheme is proposed for multiuser mmWave massive MIMO systems. The main idea is designing hybrid precoding in the digital and analog domains separately to achieve the near-optimal achievable sum rate. In the analog domain, we exploit a DFT codebook-based analog precoder. The core of the analog precoder design is the proposed conflicting-aware (CA) beam-column selection from the DFT codebook, which is from the analysis of sum rate maximum and the utilization of angular channel matrix. Compared to exhaustive search, the CA selection method dramatically reduces computational complexity. Then, in the digital domain, the digital precoder is obtained by applying ZF criterion in the effective channel matrix. After our proposed hybrid precoding, the achievable sum rate of the system can be maximized. The SR and EE performance of the schemes under different

configurations are shown from various simulation results. Our proposed algorithm with low computational complexity approaches to the full digital performance and gets better performance compared to other existing hybrid algorithms.

6 Appendix

In this section, we introduce a selection criterion of the dominant beam for the u^{th} user.

Different from retaining most of the power in previous beam selection, we use a metric T , defined by the signal to leakage plus noise ratio (SLNR) to maximize SR. SLNR for the u^{th} user is defined as

$$SLNR_u = \frac{\|\mathbf{H}_u^a \mathbf{w}_u\|_F^2}{\sum_{i=1, i \neq u}^{N_{BS}} \|\mathbf{H}_i^a \mathbf{w}_u\|_F^2 + \sigma^2}, u = 1, 2, \dots, U, i = 1, 2, \dots, N_{BS}, \quad (29)$$

where \mathbf{H}_u^a is the $N_{MS} \times N_{BS}$ matrix which indicates the angular channel between the BS and the u^{th} user, and \mathbf{w}_u is the hybrid precoding matrix described as $\mathbf{F}_{RF} \mathbf{f}_{BB}^u$.

If the leakage channel matrix for the u^{th} user is defined as $\mathbf{H}_{ul}^a = [(\mathbf{H}_1^a)^H \dots (\mathbf{H}_{u-1}^a)^H, \dots, (\mathbf{H}_{u+1}^a)^H \dots (\mathbf{H}_{N_{BS}}^a)^H]^H$, the formula (29) can be equivalently written as

$$SLNR_u = \frac{\mathbf{w}_u^H (\mathbf{H}_u^a)^H \mathbf{H}_u^a \mathbf{w}_u}{\mathbf{w}_u^H (\sigma^2 \mathbf{I} + (\mathbf{H}_{ul}^a)^H \mathbf{H}_{ul}^a) \mathbf{w}_u}. \quad (30)$$

(30) is typical generalized Rayleigh ratio for the variable \mathbf{w}_u . We can obtain that

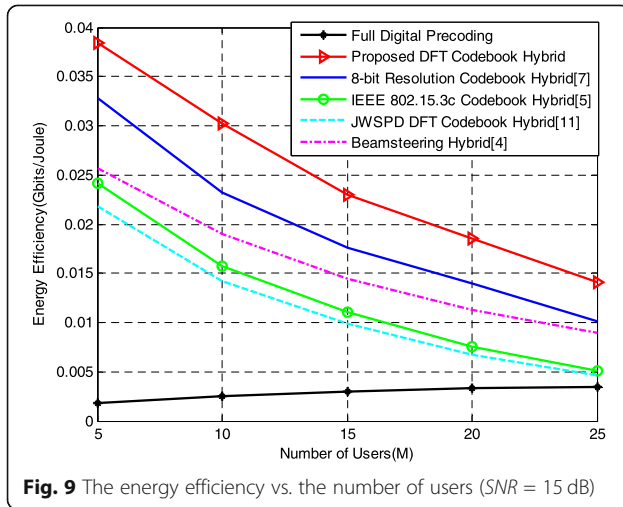
$$SLNR_u \leq \lambda_{\max} \left((\sigma^2 \mathbf{I} + (\mathbf{H}_{ul}^a)^H \mathbf{H}_{ul}^a)^{-1} (\mathbf{H}_u^a)^H \mathbf{H}_u^a \right), \quad (31)$$

where λ_{\max} indicates the maximum eigenvalue of a matrix. The maximum value of SLNR is marked as $T_{u,i}$

$$= \frac{|\mathbf{H}^a(u,i)|^2}{\sum_{j=1, j \neq i}^{N_{BS}} |\mathbf{H}^a(u,j)|^2 + \sigma^2}, \text{ which can be the metric for}$$

dominant beam selection.

The physical significance of the high SLNR value is that the equivalent channel gain of the user increases, while the power leaked to other users is still small, which means the interference to other users is small. When each user leaks to other users with less power, the interference to other users is relatively smaller. Hence, the SINR of each user can be large. Then, the achievable sum rate is high according to (14). Following this selection criterion of beams for the u^{th} user, the dominant beams make a direct contribution to SR.



Abbreviations

5G: The fifth generation of mobile telecommunications; AoA: Angle of arrival; AoD: Angle of departure; BS: Base station; CA: Conflicting aware; CSI: Channel state information; CUs: Conflicting users; DFT: Discrete Fourier transform; EE: Energy efficiency; EE: Energy efficiency; JWSPD: Joint codeword selection and precoder design; MIMO: Multiple-input multiple-output; mmWave: Millimeter wave; MS: Mobile station; MUB: Unbiased bases combined; NCUs: Non-conflicting users; RF: Radiofrequency; SINR: Signal to interference plus noise ratio; SLNR: Signal to leakage plus noise ratio; SNR: Signal to noise ratio; SR: Sum rate; SR: Sum rate; ULA: Uniform linear array; WLAN: Wireless local area networks; ZF: Zero-forcing

Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions that helped improve the quality of this manuscript.

Authors' contributions

All authors have contributed equally. All authors have read and approved the final manuscript.

Funding

This work was supported by the Natural Science Foundation of China under grant no. 61372126 and no. 61771257, Jiangsu Higher Education Institutions under grant no. 18KJB510027, Technology Projects of Jiangsu Quality and Technical Supervision grant no. KJ175943, and Postgraduate Research & Practice Innovation Program of Jiangsu Province grant no. KYCX19_0928.

Availability of data and materials

Not available online. Please contact the author for data requests.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹College of Electronic and Optical Engineering & College of Microelectronics, Nanjing University of Posts and Telecommunications, Nanjing, China.

²National and Local Joint Engineering Laboratory of RF Integration and Micro-Assembly Technology, Nanjing University of Posts and Telecommunications, Nanjing, China. ³National Quality Supervision and Testing Centre for RFID Products (Jiangsu), Jiangsu Institute of Quality and Standardization, Nanjing, China.

Received: 12 June 2019 Accepted: 21 February 2020

Published online: 05 March 2020

References

1. S. Han, C.-L. I, Z. Xu, et al., Large-scale antenna systems with hybrid precoding analog and digital beamforming for millimeter wave 5G. *IEEE Commun. Mag.* **53**(1), 186–194 (2015)
2. P. Amadori, C. Masouros, Low RF-complexity millimeter-wave beam-space MIMO systems by beam selection. *IEEE Trans. Commun.* **63**(6), 2212–2222 (2015)
3. O. El Ayach, S. Rajagopal, S. Abu-Surra, et al., Spatially sparse precoding in millimeter wave MIMO systems. *IEEE Trans. Wirel. Commun.* **13**(3), 1499–1513 (2014)
4. A. Alkhateeb, G. Leus, R.W. Heath, Limited feedback hybrid precoding for multi-user millimeter wave systems. *IEEE Trans. Wirel. Commun.* **14**(11), 6481–6494 (2015)
5. Part 15.3: wireless MAC and PHY specifications for high rate WPANs. Amendment 2: millimeter-wave-based alternative physical layer extension, IEEE Std 802.15.3c-2009. (2009).
6. Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications. Amendment 3: enhancements for very high throughput in the 60 GHz Band, IEEE Std. 802.11ad-2012. (2012).
7. S. Kuttu, D. Sen, Beamforming for millimeter wave communications: an inclusive survey. *IEEE Commun. Surv. Tutorials* **18**(2), 949–973 (2016)
8. L. Zhou, Y. Ohashi, *Efficient codebook-based MIMO beamforming for millimeter-wave WLANs* (IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Commun. (PIMRC), Sydney, 2012), pp. 1885–1889
9. K. Satyanarayana, M. El-Hajjar, et al., *Millimeter wave hybrid beamforming with DFT-MUB aided precoder codebook design* (2017 IEEE 86th Vehicular Technology Conference (VTC-Fall), Toronto, 2017)
10. Y. Han, S. Jin, J. Zhang, et al., DFT-based hybrid beamforming multiuser systems: rate analysis and beam selection. *IEEE J. Select Top. Signal Process.* **12**(3), 514–528 (2018)
11. S. He, J. Wang, Y. Huang, et al., Codebook-based hybrid precoding for millimeter wave multiuser systems. *IEEE Trans. Signal Process.* **65**(20), 5289–5304 (2017)
12. T.M. Pham, R. Farrell, et al., Efficient zero-forcing precoder design for weighted sum-rate maximization with per-antenna power constraint. *IEEE Trans. Vehicular Technol.* **67**(4), 3640–3645 (2017)
13. G.H. Golub, C.F. Van Loan, *Matrix computations*, 4th edn. (The Johns Hopkins University Press, Baltimore, 2013), p. 89
14. T.S. Rappaport, F. Gutierrez, et al., Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications. *IEEE Trans. Antennas Propagation* **61**(4), 1850–1859 (2013)
15. I.A. Hemadeh, K. Satyanarayana, et al., Millimeter-wave communications: physical channel models, design considerations, antenna constructions, and link-budget. *IEEE Commun. Surv. Tutorials* **20**(2), 870–913 (2018)
16. T.S. Rappaport, S. Sun, R. Mayzus, et al., Millimeter wave mobile communications for 5G cellular: it will work! *IEEE Access* **1**, 335–349 (2013)
17. D. Tse, P. Viswanath, *Fundamentals of wireless communication* (Cambridge University Press, Cambridge, 2005)
18. Y. Zeng, R. Zhang, Millimeter wave MIMO with lens antenna array: a new path division multiplexing paradigm. *IEEE Trans. Commun.* **64**(4), 1557–1571 (2016)
19. C.-S. Lee, W.-H. Chung, *Max-min hybrid precoding in millimeter wave cooperative MISO systems* (Conf. IEEE International Conference on Communications (ICC), Kuala Lumpur, 2016), pp. 1–6
20. K. Satyanarayana, M. El-Hajjar, P.-H. Kuo, A.A.M. Mourad, L. Hanzo, Adaptive transceiver design for C-RAN in mmWave communications. *IEEE Access* **6**, 16770–16782 (2017)
21. S. Sanayei, A. Nosratinia, Antenna selection in MIMO systems. *IEEE Commun. Mag.* **42**(10), 68–73 (2004)
22. H. Moon, Waterfilling power allocation at high snr regimes. *IEEE Trans. Commun.* **59**(3), 708–715 (2011)
23. X. Gao, L. Dai, S. Han, C.-L. I, R.W. Heath, Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays. *IEEE J. Sel. Areas Commun.* **34**(4), 998–1009 (2016)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)