

RESEARCH

Open Access



Spectral-efficient hybrid precoding for multi-antenna multi-user mmWave massive MIMO systems with low complexity

Yang Liu^{1†}, Qitong Zhang^{1†}, Xin He¹, Xuemei Lei^{1*} , Yinghui Zhang¹ and Tianshuang Qiu²

[†]Yang Liu and Qitong Zhang contributed equally to this work

*Correspondence: xmleimu@163.com

¹ College of Electronic Information Engineering, Inner Mongolia University, Hohhot 010021, China

² Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024, China

Abstract

Millimeter-wave (mmWave) massive multiple-input multiple-output (MIMO) systems allow for a data transmission rate of gigabits per second owing to the large bandwidth available in the mmWave spectrum and the antenna gains provided by the massive MIMO system. However, hybrid precoding with high complexity and low spectral efficiency cannot address the challenge of high cost and power consumption of RF chains of multi-user systems. In this paper, we propose a low-complexity hybrid precoding scheme for downlink multi-antenna multi-user mmWave massive MIMO systems, aiming to enhance the sum spectral efficiency (SSE) performance. We first extend the dimension of the analog precoding matrix into a square matrix and find the optimal analog combiner by selecting some of the discrete Fourier transform (DFT) bases, which enhances the equivalent baseband channel matrix gain. Then, we directly aggregate the channel gain through the equal gain transmission (EGT) method to ensure the frequency efficiency performance. Finally, we propose an improved BD scheme to design the digital precoder and combiner to reduce the inter-user interference. We consider both the mmWave channel and the Rayleigh channel to evaluate the performance of the proposed algorithm. The simulation results verify that the proposed scheme enjoys near-optimal achievable sum spectrum efficiency and BER performance in both the mmWave channel and Rayleigh channel and performs even better in Rayleigh channel than in the mmWave channel.

Keywords: Millimeter wave, Massive MIMO, Hybrid precoding, Block diagonalization, Spectral efficiency

1 Introduction

The rapid increase in mobile data growth and the diversification of the application scenarios of 5 G and beyond are creating unprecedented challenges for wireless service providers in overcoming a global bandwidth shortage. This shortage has motivated the exploration of the millimeter-wave (mmWave) frequency spectrum, which can provide hundreds of free GHz spectrum resources for future cellular communication networks [1]. However, mmWave easily leads to a high propagation loss due to its very small wavelength [2]. To solve this problem, researchers have combined mmWave and massive multiple-input multiple-output (MIMO) techniques to increase the capacity of the

communication systems [3]. A massive MIMO antenna array can overcome high path loss in the mmWave channel. Furthermore, the small wavelength and narrow beam characteristics of millimeter waves allow systems to employ large-scale antenna arrays [4], which can be packed into small-form factors and achieve reasonable array gains. Therefore, the combination of mmWave and massive MIMO systems can effectively realize the complementary advantages of the two technologies [5–8].

Implementing high-quality communication links in mmWave massive MIMO systems requires the deployment of large antenna arrays at base station (BS). Each BS needs to serve multiple mobile users (MU) simultaneously for efficient system performance. Therefore, precoding is applied to generate the transmitted signal at the BS for multiplexing various data streams to different users and canceling out noise, inter-user interference and fast fading in the mmWave massive MIMO systems. Traditional full digital precoding reduces the interference between data streams and users and the complexity of receiver computation [9, 10]. However, full digital precoding schemes require every antenna element to be driven by its own dedicated energy-intensive radio frequency (RF) chain, imposing extremely high costs and energy consumption [11]. In contrast, an antenna array of full analog precoding is connected to only one RF chain by a phase shifter [12, 13]. Since the phase shifter cannot control the amplitude, the performance of full analog precoding scheme is restricted. To solve these problems, hybrid precoding schemes have been proposed that implement cost-effective variable phase shifters in the RF domain and allow reduced-dimensional signal processing schemes to be carried out digitally in the baseband [14]. Hybrid precoding can reduce the number of RF chains without obvious performance loss [15], thus, achieving a tradeoff between system performance and hardware complexity. Because of the characteristics of anti-multipath fading [16], anti-interference, high spectral efficiency [17], high energy efficiency [18], etc., hybrid precoding has become an important signal processing technology in mmWave massive MIMO systems [19–22].

Hybrid precoding algorithms in mmWave massive MIMO system can be divided into three categories according to the number of users and the number of antennas. The first is the single-user system in which the user carries multiple antennas. [23] proposed a hybrid precoding that selects analog and digital precoder from a discrete Fourier transform (DFT)-based codebook to maximize the spectral efficiency in single-user system. The theoretical basis of this algorithm is to minimize the chordal distance between the optimal unconstrained precoder obtained from the maximum right singular vector of the channel and the hybrid beamformers selected from the statistical DFT codebook. However, it aims to minimize the performance loss under limited feedback CSI under transmitters assumptions. [24] proposed a double-pilot-based hybrid precoding system, which predicts the analog precoding by using deep learning method and updates the equivalent channel frequently for the digital precoding by enhancing the frequency of equivalent channel estimation. This method is flexible but with high consumption of pilot and computation time because the time-varying influence. [25] proposed a hybrid precoding algorithm based on switches that selects a subset of antennas to feed through the RF switching network and delivers a lower-dimensional array to the digital domain. Although this algorithm permits low-complexity and hardware implementation, devising such a hybrid beamformer can be computationally expensive for large-dimensional

systems due to the binary nature of the selection problem. Since the single-user system cannot meet the practical needs that the system simultaneously serves multiple user devices, multi-user systems has been attracted more attention. There are two categories in multi-user systems: One is that each user carries a single antenna, and the other is that each user adopts multiple antennas to transmit signals. In single-antenna multi-user system, it is difficult to jointly consider the changes of user-specific signal quality and multi-user interference. [26] proposed a low-cost hybrid precoding algorithm based on quantized CSI feedback for a single-antenna scenario and revealed that pure analog precoding outperforms hybrid precoding in terms of the ergodic achievable rate under certain conditions, as derived in closed forms with respect to the SNR, the number of users and the number of feedback bits. However, directly quantized feedback still causes a huge amount of overhead to maintain a usable accuracy. In [27], a deep learning quantized phase hybrid precoding algorithm was proposed. This algorithm estimates the channel vectors with a deep compression sensing algorithm and trains the hybrid precoding neural network with the estimated channel vectors in the offline state. Then, the analog precoding matrix can be obtained by ideal phase quantization and the digital precoding matrix can be obtained via ZF algorithm in the online state. However, the performance of deep learning-based CSI feedback architecture will decline significantly when the dataset is insufficient. Furthermore, the compressed sensing method used to obtain the estimated channel matrix is iterative, which will also cause poor real-time performance.

Since single-antenna multi-user systems cannot meet the requirements of multiple antenna configurations for user terminals, multi-antenna multi-user systems with larger antenna array and stronger interference have been attracted more attention in recent years. Hybrid precoding in multi-antenna multi-user mmWave massive MIMO system can be divided into joint optimization and two-stage optimization for analog precoding and digital precoding according to the design steps. In [28], a convolutional neural network (CNN) framework was first proposed to optimize hybrid precoding for mmWave multi-user MIMO (MU-MIMO) systems in which the network takes the imperfect channel matrix as the input and jointly produces analog and digital precoders as outputs. This method can effectively enhance the performance and efficiency of the system, but using multiple large-dimensional layers constructions may consume tremendous computation time in the training phase, which is impractical with the hardware constraints (e.g., limited computational capability and memory resources) of mobile terminals. Generally, the joint optimization algorithm should be an essential part of designing a precoding scheme because the comprehensiveness of problem formulation. Nevertheless, the signaling overhead is very heavy because the joint optimization requires CSI for the whole system which is equipped with large dimension antenna arrays. Additionally, the optimal solution is hardly tractable because the jointly optimization is a non-convex mixed integer optimization. Moreover, the digital precoding and analog beamforming are asynchronous process in practical application. Therefore, to eliminate the hardware restrictions and consider the asynchronism of analog and digital precoding process, the two-stage algorithms are used in practical design [29–32]. Although the two-stage method may not achieve optimal precoding design compared to the joint optimization with comprehensiveness of problem

formulation, the solution of the two-stage method needs less feedback and training signals for being considered as a non-synchronous process. Therefore, the two-stage method has been widely appreciated in recent years. [29] proposed an alternating optimization scheme to design the analog precoder and combiner, and [30] sorted the antenna sub-arrays according to the capacity of the channel before optimization. The above two algorithms adopt the iterative optimization theory, which increases the computational complexity of implementation. In [31], the author developed an adaptive two-stage reduced dimensionality multi-user hybrid precoder algorithm with limited feedback. The considered model assumes that each user only supports a single data stream. This problem becomes more complicated if the transmission of multiple data streams per user and the hybrid structure implemented by the user are also taken into account. [32] proposed a two-stage hybrid precoding algorithm based on equal gain transmission (EGT) aiming at maximizing the end-to-end mutual information (EEMI). Apart from the low complexity of its implementation, the developed scheme is a more generic hybrid block diagonalization (HBD) solution, as it not only takes the frequency selectivity into account but also removes the reliance on the high-resolution analog network. However, the traditional EGT method regards the problem of finding optimal analog precoding/combiner matrix as a multivariate coupling variable optimization that has non-convex nonlinearity. Moreover, the spectral efficiency performance of this algorithm was degraded in the case of serious interference between users. To achieve a tradeoff between system performance and complexity, two-stage hybrid precoding with low complexity and high spectral efficiency becomes a crucial solution to combat path loss and interference in multi-antenna multi-user mmWave massive MIMO system.

In this paper, we propose a spectral-efficient hybrid precoding algorithm for a down-link multi-antenna multi-user mmWave massive MIMO system. To reduce the hardware restriction and iterative process of the algorithm while ensuring the performance, we split the design into analog precoder/combiner and the digital counterparts. Because the large spectral efficiency can be obtained by increasing the equivalent baseband channel gain, we first propose to extend the dimension of the analog precoding matrix into a full rank matrix and find the optimal analog combiner in analog precoder/combiner processing. Then, we perform the analog combiner based on an EGT method to aggregate the channel gain. Different from traditional EGT algorithms [32], we propose to represent the codebook by a discrete Fourier transform (DFT) matrix because each column vector in DFT matrix is unrelated and the column vectors can be combined linearly to synthesize the array response vector in any direction [33], which makes the analog combiner design to be a sparse precoding optimization problem containing only unary variables. Moreover, because the number of receiving antennas per user in mmWave massive MIMO is much smaller than the number of BS antennas and the hardware restrictions in practical applications, the exhaustive search on the DFT bases is acceptable. Finally, BD precoding is performed based on the equivalent channel matrix to eliminate the inter-user interference. The contributions of the paper are summarized as follows.

- (1) We divide the calculation of hybrid precoding into analog and digital stages to reduce the signaling overhead and formulate the spectral efficiency optimization

as maximizing the equivalent channel gain in multi-antenna multi-user mmWave massive MIMO system.

- (2) We propose to denote the analog precoder as a square matrix to obtain a large equivalent channel gain and use the EGT method to reap the diversity benefit of an analog phased-array. Then, we construct a DFT basis by discretizing the codebook which is extracted from the equivalent channel gain, so that we can select vectors from the DFT basis for analog combiner design. The proposed design can obtain a large array gain by enlarging the sum squares of diagonal terms with low computational complexity.
- (3) Furthermore, we use a low-dimensional block diagonalization digital precoding and combining to attenuate the complexity and significantly maximize the spectral efficiency by reducing the inter-user interference of the multi-antenna multi-user system.
- (4) The theoretical analysis shows that the proposed hybrid precoding scheme can improve the spectral efficiency with low complexity. We also provide numerical results to demonstrate that the proposed scheme can achieve a better spectral efficiency and bit-error-rate than some existing schemes under both Saleh–Valenzuela mmWave and Rayleigh channels.

The rest of this paper is organized as follows: In Sect. 3, the system model for a multi-user massive MIMO system is described. In Sect. 4, a low-complexity hybrid precoding algorithm is proposed for downlink multi-antenna multi-user mmWave massive MIMO systems, aiming to enhance the sum spectral efficiency (SSE) performance. Moreover, the computational complexity of the proposed algorithm and other traditional algorithms are compared in a theoretical analysis. The simulation results and the conclusion are presented in Sects. 5 and 6, respectively.

Notation: We use bold-faced lower-case and upper-case letters to denote column vector and Rayleigh matrices, respectively. $(\cdot)^{-1}$ and $(\cdot)^H$ represent the inversion and conjugate of a matrix, respectively. $E[\cdot]$ denotes the expectation value. $\|\cdot\|$ and $\|\cdot\|_F$ denote the norm of a vector and the Frobenius norm of a matrix, respectively. Finally, $\mathcal{CN}(0, 1)$ represents the complex Gaussian distribution with zero mean and unit variance.

2 Methodology

In this paper, we first introduce the research background and related hybrid precoding methods for mmWave massive MIMO systems in Sect. 1. Compared with single-user scenario, multi-user scenario satisfies the practical demand for multi-user service in modern communication systems. Compared with multi-user single-antenna scenario, the multi-user multi-antenna scenario meets the number requirements of cells which needed to be served in the actual mmWave communication system. However, existing hybrid precoding algorithms cannot maximize their spectral efficiency with low computational complexity because the antenna array is more complex and there is more interference in multi-user multi-antenna scenarios. Therefore, hybrid precoding with low complexity and high spectral efficiency becomes the key solution to combat interference and path loss in mmWave massive MIMO systems.

We propose a spectral-efficient hybrid precoding algorithm in a downlink multi-user multiple-antenna mmWave massive MIMO system to maximize the sum spectral efficiency and reduce the complexity of the system. To reduce the iterative process of the algorithm while ensuring the performance, we split the precoding design into the RF precoding and the digital precoding design. In RF precoding, we extend the dimension of the analog precoding matrix into a square matrix and find the optimal analog combiner by selecting some of the discrete Fourier transform bases, which enhances the equivalent baseband channel matrix gain. Then, to ensure the spectral efficiency performance, we directly aggregate the channel gain through the equal gain transmission method. In terms of optimizing digital precoding, BD technology is used under equivalent channels to eliminate the inter-user interference.

To verify the effectiveness of the proposed algorithm, we have conducted a variety of experiments to obtain comparison results by calculating the computational complexity and simulating in the same environment in both mmWave and Rayleigh channel. The specific analysis can be found in Sect. 5.

3 System model

3.1 Multi-user massive MIMO system model

A downlink transmission of a TDD-based multi-antenna massive MU-MIMO system is depicted in Fig. 1. The base station (BS) is equipped with N_t transmission antennas and N_{RF}^t RF chains to communicate with K independent users. Each user is equipped with N_r receiving antennas and N_{RF}^r RF chains to support N_s data streams. To ensure the effectiveness of transmission, the number of transmitted steams is constrained by $KN_s \leq N_{RF}^t \leq N_t$ for the BS and $N_s \leq N_{RF}^r \leq N_r$ for each user.

The transmitted symbols are processed by a $N_{RF}^t \times KN_s$ digital precoder F_{BBk} and mapped onto N_{RF}^t RF chains. Then, the symbols are processed by a $N_t \times N_{RF}^t$ analog precoder F_{RF} . Since the analog precoder is composed of analog phase shifters, it can only control the phase of the signal and each entry of F_{RF} is normalized to satisfy $|F_{i,j}| = \frac{1}{\sqrt{N_t}}$, where $|F_{i,j}|$ denotes the amplitude of the (i, j) -th element of F_{RF} . The digital precoder F_{BBk} enables both amplitude and phase modifications.

We assume $H_k \in C^{N_t \times N_r}$ as the downlink channel matrix for the k -th user. The received signal of the k -th user can be expressed as

$$y_k = H_k F_{RF} F_{BBk} s_k + n_k, k = 1, 2 \dots, K, \tag{1}$$

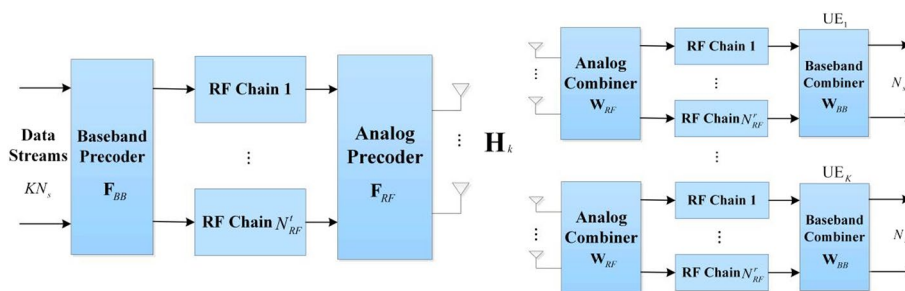


Fig. 1 System model of multi-user hybrid precoding

where $\mathbf{S} = [\mathbf{s}_1^T, \mathbf{s}_2^T, \dots, \mathbf{s}_K^T]^T$ is the signal vector of K users, $\mathbf{S} \in \mathbb{C}^{KN_s \times 1}$ satisfies $E[\mathbf{S}\mathbf{S}^H] = \frac{P}{KN_s} \mathbf{I}_{KN_s}$, P represents the average transmit power of the BS, and \mathbf{I}_{KN_s} is a $KN_s \times KN_s$ identity matrix. $\mathbf{n}_k \in \mathbb{C}^{N_r \times 1}$ is the vector of the i.i.d. $\mathcal{CN}(0, 1)$ additive white Gaussian noise (AWGN). The received signal \mathbf{y}_k after combining at the k -th user can be represented as

$$\hat{\mathbf{y}}_k = \mathbf{W}_k^H \mathbf{W}_{RFk}^H \mathbf{H}_k \mathbf{F}_{RF} \mathbf{F}_{BBk} \mathbf{s}_k + \mathbf{W}_{BBk}^H \mathbf{W}_{RFk}^H \mathbf{n}_k, k = 1, 2, \dots, K, \tag{2}$$

where \mathbf{W}_{RFk} is the $N_r \times N_{RF}^r$ analog combiner matrix and \mathbf{W}_{BBk} is the $N_{RF}^t \times N_s$ baseband combiner for the k -th user.

If $\tilde{\mathbf{H}}_k = \mathbf{W}_{RFk}^H \mathbf{H}_k \mathbf{F}_{RF}$, $k = 1, 2, \dots, K$ is denoted as the equivalent baseband channel matrix, then (2) can be re-expressed as

$$\hat{\mathbf{y}}_k = \mathbf{W}_{BBk}^H \tilde{\mathbf{H}}_k \mathbf{F}_{BBk} \mathbf{s}_k + \sum_{i=1, i \neq k}^K \mathbf{W}_{BBk}^H \tilde{\mathbf{H}}_i \mathbf{F}_{BBi} \mathbf{s}_i + \mathbf{W}_{BBk}^H \mathbf{W}_{RFk}^H \mathbf{n}_k, k = 1, 2, \dots, K. \tag{3}$$

Therefore, the sum spectral efficiency of K users can be expressed as

$$\mathbf{R} = \sum_{k=1}^K \log_2 \left(\left| I_{N_s} + \frac{P}{KN_s} \mathbf{R}_i^{-1} \mathbf{W}_{BBk}^H \tilde{\mathbf{H}}_k \mathbf{F}_{BBk} \mathbf{F}_{BBk}^H \tilde{\mathbf{H}}_k^H \mathbf{W}_{BBk} \right| \right), \tag{4}$$

where $\mathbf{R}_i = \frac{P}{KN_s} \sum_{i=1, i \neq k}^K \mathbf{W}_{BBk}^H \tilde{\mathbf{H}}_k \mathbf{W}_{BBk} + \sigma^2 \mathbf{W}_{BBk}^H \mathbf{W}_{RFk}^H \mathbf{W}_{RF} \mathbf{W}_{BBk}$ is the covariance matrix of interference and noise.

3.2 Channel model

In this paper, the general channel matrix is set as $\mathbf{H} = [\mathbf{H}_1^T, \mathbf{H}_2^T, \dots, \mathbf{H}_K^T]^T = [\sqrt{\beta_1} \dot{\mathbf{H}}_1^T, \sqrt{\beta_2} \dot{\mathbf{H}}_2^T, \dots, \sqrt{\beta_K} \dot{\mathbf{H}}_K^T]^T$, where $\sqrt{\beta_k}$ indicates large-scale path fading and $\dot{\mathbf{H}}_k$ indicates the normalized channel matrix for the k -th user, satisfying $E[\|\dot{\mathbf{H}}_k\|_F^2] = N_t N_r$. In the mmWave channel, the main feature is the limited number of scattering clusters in the propagation path [34–36]. To characterize the limited scattering property of mmWave channels, the Saleh–Valenzuela geometric model is adopted in this paper [37, 38]. Specifically, the channel $\mathbf{H}_k \in \mathbb{C}^{N_t \times N_r}$ from the BS to the k -th user can be modeled as

$$\mathbf{H}_k = \sqrt{\frac{N_t N_r}{N_{cl} N_{ray}}} \sum_{u=1}^{N_{cl}} \sum_{l=1}^{N_{ray}} \alpha_{ul}^k \mathbf{a}_r^k(\theta_{ul}^r) \mathbf{a}_t^k(\phi_{ul}^t)^H, \tag{5}$$

where N_{cl} is the number of clusters and each cluster comprises N_{ray} propagation paths. α_{ul}^k is the complex amplitude associated with the l -th path in the u -th scattering cluster. $\mathbf{a}_r^k(\theta_{ul}^r)$ and $\mathbf{a}_t^k(\phi_{ul}^t)$ represent the receive and transmit array response vectors for the k -th user, respectively, whereas θ_{ul}^r and ϕ_{ul}^t are the arrival and departure azimuth angles (AOA and DOA) of the (u, l) -th path. The truncated Laplacian distribution is employed to generate θ_{ul}^r and ϕ_{ul}^t [39]. Although the proposed algorithms can be implemented on any antenna array, we use a uniform linear array (ULA) as the array model, the array response vector can be defined as

$$\mathbf{a}_{ULA}(\theta) = \frac{1}{\sqrt{N}} \left[1, e^{j\frac{2\pi}{\lambda}d \sin(\theta)}, \dots, e^{j(N-1)\frac{2\pi}{\lambda}d \sin(\theta)} \right]^T, \tag{6}$$

where λ is the wavelength of the signal, and d is the distance between adjacent antenna elements. It is always assumed that $d = \frac{\lambda}{2}$.

4 Proposed hybrid precoding algorithm

Generally, the joint optimization of the digital hybrid precoder and analog combiner should be an essential part of designing a processing scheme for optimal sum spectral efficiency [28, 40]. However, the optimal solution is hardly tractable because the jointly optimization is a non-convex mixed integer optimization. Moreover, the digital precoding and analog precoding are asynchronous process in practical application. Therefore, to eliminate the hardware restrictions and consider the asynchronism of analog and digital precoding process, we divide the calculation of the hybrid precoding into analog and digital steps [29–32]. In the first step, the BS and each MS design the analog precoding and combining vectors to maximize the desired array gain for equivalent channel matrix. In the second step, the digital precoder is designed based on the equivalent channel matrix after the application of the proposed analog precoder in the first stage, to manage the multi-user interference. In general, the two-stage method is an asynchronous calculation of analog precoding and digital precoding based on the equivalent channel matrix.

4.1 The analog precoder/combiner design

Owing to the large number of antennas in massive MU-MIMO systems, the channel gains of the equivalent channel \mathbf{H}_{eq} can be scaled up through appropriate phase-only control in the RF domain. In traditional multi-user hybrid precoding algorithms [41, 42], because the dimension of the analog precoding matrix \mathbf{F}_{RF} is $N_t \times N_{RF}^t$, the rank of the baseband equivalent channel matrix \mathbf{H}_{eq} is reduced after passing through the channel. This result is a reduction in the gain of the baseband equivalent channel matrix, thereby reducing the sum spectral efficiency. To obtain a large array gain, it is necessary to increase the rank of \mathbf{F}_{RF} . We first denote the analog precoding matrix as an $N_t \times N_t$ full rank matrix. Thus, the improved equipment baseband multi-user channel \mathbf{H}_{eq} can be expressed as

$$\mathbf{H}_{eq} = \begin{bmatrix} \tilde{\mathbf{H}}_1 \\ \tilde{\mathbf{H}}_2 \\ \vdots \\ \tilde{\mathbf{H}}_K \end{bmatrix} = \begin{bmatrix} \mathbf{W}_{RF1}^H & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_{RF2}^H & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{W}_{RFK}^H \end{bmatrix} \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \\ \vdots \\ \mathbf{H}_K \end{bmatrix} \mathbf{F}_{RF}. \tag{7}$$

To reap the diversity benefit of an analog phased-array, we utilize the analog precoding based on EGT method. The analog precoding matrix can be expressed as

$$\mathbf{F}_{i,j} = \frac{1}{\sqrt{N_t}} e^{j\varphi_{ij}}, \tag{8}$$

where $\varphi^{(ij)}$ is the phase of the (i, j) -th element of $(\mathbf{W}_{RF}\mathbf{H})^H$.

Since the diagonal elements in \mathbf{H}_{eq} represent the equivalent channel gain, the off-diagonal elements in \mathbf{H}_{eq} indicate the inter-chain interference. Therefore, to obtain

the large array gain, we propose to increase the sum of the squares of the diagonal elements in \mathbf{H}_{eq} by finding the precoding combiners \mathbf{W}_{RFk} ($k = 1, 2 \dots K$) of K users. The optimization problem is described in

$$\begin{aligned} \max_{\mathbf{W}_{RFk}} & \sum_{k=1}^K \sum_{m=1}^{N_{RF}^r} \left\| \left(\mathbf{w}_{RFk}^{(m)} \right)^H \mathbf{H}_k \right\|_1^2, \\ \text{s.t.} & \left| \mathbf{W}_{RFk}^{(i,j)} \right| = \frac{1}{\sqrt{N_{RF}^r}}, \forall i, j \end{aligned} \tag{9}$$

where $\mathbf{w}_{RFk}^{(m)}$ is the m -th column of \mathbf{W}_{RFk} , $\left\| \left(\mathbf{w}_k^{(m)} \right)^H \mathbf{H}_k \right\|_1$ is the $((k-1)N_{RF}^r + m)$ -th diagonal element of \mathbf{H}_{eq} and $\|\cdot\|_1$ denotes the 1-norm of a vector, corresponding to the m -th RF chains of the k -th user. Because all \mathbf{W}_{RFk} ($k = 1, 2 \dots K$) are independent of each other, the above equation is equivalent to maximizing (9) for all K users. Then, the analog combiner can be obtained from the following optimization,

$$\begin{aligned} \max_{\mathbf{W}_{RFk}} & \sum_{m=1}^{N_{RF}^r} \left\| \left(\mathbf{w}_{RFk}^{(m)} \right)^H \mathbf{H}_k \right\|_1^2, k = 1, 2 \dots K \\ \text{s.t.} & \left| \mathbf{W}_{RFk}^{(i,j)} \right| = \frac{1}{\sqrt{N_{RF}^r}}, \forall i, j \end{aligned} \tag{10}$$

Due to the non-convex constraints of \mathbf{W}_{RFk} , we cannot directly solve the problem in (9). Therefore, we modify the constraints of the analog combining matrix by

$$\left\| \left(\mathbf{w}_{RFk}^{(m)} \right)^H \mathbf{H}_k \right\|_1^2 = \left(\sum_{n=1}^{N_t} \left| \left(\mathbf{w}_{RFk}^{(m)} \right)^H \mathbf{h}_k^{(n)} \right| \right)^2, \tag{11}$$

where $\mathbf{h}_k^{(n)}$ denotes the n -th column of \mathbf{H}_k . Moreover, in the geometric MIMO channel models, including the Rayleigh fading and mmWave channels, $\mathbf{h}_k^{(n)}$ is the linear combination of the array response vectors of all AOAs. That is, each addition term $\left| \left(\mathbf{w}_{RFk}^{(m)} \right)^H \mathbf{h}_k^{(n)} \right|$ in $\left\| \left(\mathbf{w}_{RFk}^{(m)} \right)^H \mathbf{H}_k \right\|_1$ is the absolute value of the weighted sum of the array response vectors $\mathbf{a}_t^k(\phi_{il}^t)$ of all AoAs projected onto $\mathbf{w}_{RFk}^{(m)}$. Therefore, we can set $\mathbf{w}_{RFk}^{(m)}$ in the form of an array response vector (6) to extract the gain from projections,

$$\mathbf{c}(w) = \frac{1}{\sqrt{N_r}} \left[1, e^{jw}, e^{j2w}, \dots, e^{j(N_r-1)w} \right]^T, \tag{12}$$

where $w = \frac{2\pi}{\lambda} d \sin \theta$ represents the corresponding spatial frequency.

To obtain the large array gain, we require that the rank of the baseband equivalent channel matrix \mathbf{H}_{eq} does not decrease after the analog combining matrix \mathbf{W}_{RFk} is multiplied by the channel matrix \mathbf{H}_k . Therefore, we require the columns of \mathbf{W}_{RFk} to be pairwise orthogonal so that the rank of \mathbf{H}_{eq} is lower bound by $N_{RF}^r > N_s$. Traditional EGT algorithms, such as [43, 44], find the optimal analog combiner via iterative methods, which has non-convex nonlinearity and is time-consuming and computationally complex. Considering that each column (representing the antenna response in each incidence direction) of DFT codebook is orthogonal to each other, we propose a DFT codebook-based analog combiner whose orthogonal beam columns are

specified by DFT codeword. The DFT base is a sparse base so that the problem formulation is transformed from multivariate coupling variable optimization problem to unitary sparse optimization problem, which reduces the computational complexity. According to the form of $\mathbf{w}_{RFk}^{(m)}$, we discretize w into N_r levels over $[0, 2\pi)$ and construct N_{RF}^r discrete Fourier transform (DFT) bases. The DFT base is a sparse base with orthogonal properties. It can be expressed as

$$\mathbf{C} = \left\{ \mathbf{c}(0), \mathbf{c}\left(\frac{2\pi}{N_{RF}^r}\right), \dots, \mathbf{c}\left(\frac{2\pi(N_{RF}^r - 1)}{N_{RF}^r}\right) \right\}, \tag{13}$$

where \mathbf{C} is a candidate set of $\mathbf{w}_{RFk}^{(m)}$ with N_{RF}^r DFT bases. Therefore, the design of the modified analog combiner can be re-expressed as

$$\begin{aligned} \max_{\mathbf{W}_{RFk}} \quad & \sum_{m=1}^{N_{RF}^r} \left\| \left(\mathbf{w}_{RFk}^{(m)} \right)^H \mathbf{H}_k \right\|_1^2 \\ \text{s.t.} \quad & \mathbf{w}_{RFk}^{(m)} \in \mathbf{C}, m = 1, 2, \dots, N_{RF}^r \end{aligned} \tag{14}$$

To solve the problem in (14), we sort all the $\left\| \left(\mathbf{w}_{RFk}^{(m)} \right)^H \mathbf{H}_k \right\|_1$ in descending order and construct the \mathbf{W}_{RFk} with the first N_{RF}^r DFT bases in $\mathbf{c}(w)$. Although various cross-domain collaborative filtering (CDCF) algorithms are presented to address the sparsity problem [45, 46], as seen from the above analysis in our method, each user needs to solve (14) only once to obtain the corresponding analog combining matrix \mathbf{W}_{RFk} . In addition, the number of receiving antennas per user should be much smaller than the number of BS antennas, which makes the exhaustive search on the DFT bases acceptable.

4.2 The digital precoder design

After obtaining the analog precoder \mathbf{F}_{RF} and combiner \mathbf{W}_{RFk} , in this section, we apply a digital BD precoding scheme to eliminate the interference among users and maximize the spectral efficiency of the system. The BD scheme is an extension of the ZF scheme in multi-user MIMO systems. Each user's linear precoder and combiner can be obtained by two singular value decomposition (SVD) operations [47]. In our hybrid case, the BD scheme is designed based on the effective channel matrix $\tilde{\mathbf{H}}_k, \forall k$. To eliminate interference among multiple users, the following constraints need to be imposed on the baseband equivalent channel matrix

$$\tilde{\mathbf{H}}_k \mathbf{F}_{BBi} = 0 (i \neq k). \tag{15}$$

It can be concluded from (15) that \mathbf{F}_{BBi} must be found in the null space of the other users' channel matrices. Thus, $\bar{\mathbf{H}}_k$ is defined as

$$\bar{\mathbf{H}}_k = \left[\tilde{\mathbf{H}}_1^T, \dots, \tilde{\mathbf{H}}_{k-1}^T, \tilde{\mathbf{H}}_{k+1}^T, \dots, \tilde{\mathbf{H}}_K^T \right]^T. \tag{16}$$

We compute the SVD of $\bar{\mathbf{H}}_k$ and obtain

$$\bar{\mathbf{H}}_k = \bar{\mathbf{U}}_k \bar{\mathbf{\Sigma}}_k \left[\bar{\mathbf{V}}_k^{(K-1)N_{RF}^r}, \bar{\mathbf{V}}_k^{N_{RF}^r} \right]^H, \tag{17}$$

where $\bar{\Sigma}_k$ is a diagonal matrix that consists of nonnegative singular values of $\bar{\mathbf{H}}_k$ and its dimension is equal to the rank of $\bar{\mathbf{H}}_k$. $\bar{\mathbf{V}}_k^{N_{RF}^r}$ and $\bar{\mathbf{V}}_k^{(K-1)N_{RF}^r}$ are composed of right singular vectors corresponding to zero singular values and nonzero singular values, respectively. Multiplying $\bar{\mathbf{V}}_k^{N_{RF}^r}$ by $\bar{\mathbf{H}}_k$, we get

$$\begin{aligned} \bar{\mathbf{H}}_k \bar{\mathbf{V}}_k^{N_{RF}^r} &= \bar{\mathbf{U}}_k \begin{bmatrix} \bar{\Sigma}_k^{((K-1)N_{RF}^r)} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{V}}_k^{(K-1)N_{RF}^r H} \\ \bar{\mathbf{V}}_k^{N_{RF}^r H} \end{bmatrix} \bar{\mathbf{V}}_k^{N_{RF}^r} \\ &= \bar{\mathbf{U}}_k \bar{\Sigma}_k^{((K-1)N_{RF}^r)} \left(\bar{\mathbf{V}}_k^{(K-1)N_{RF}^r} \right)^H \bar{\mathbf{V}}_k^{N_{RF}^r} \\ &= \mathbf{0} \end{aligned} \tag{18}$$

Thus, $\bar{\mathbf{V}}_k^{N_{RF}^r}$ is an orthogonal basis for the null space of $\bar{\mathbf{H}}_k$. This means that only the k -th user can receive the signal transmitted from the direction of N_{RF}^r . Thus, we know

$$\tilde{\mathbf{H}}_i \bar{\mathbf{V}}_k^{N_{RF}^r} = \begin{cases} \mathbf{0}, & i \neq k \\ \tilde{\mathbf{H}}_k \bar{\mathbf{V}}_k^{N_{RF}^r}, & i = k \end{cases} \tag{19}$$

Then, the digital precoder of the k -th user can be defined as $\mathbf{F}_{BBk} = \bar{\mathbf{V}}_k^{N_{RF}^r}$, and the digital precoding matrix is

$$\mathbf{F}_{BB} = \left[\bar{\mathbf{V}}_1^{N_{RF}^r}, \dots, \bar{\mathbf{V}}_K^{N_{RF}^r} \right]. \tag{20}$$

Therefore, after eliminating the inter-user interference by digital precoder \mathbf{F}_{BBi} , the equivalent baseband channel matrix \mathbf{H}_{eq} can be expressed as

$$\mathbf{H}_{BD} = \mathbf{H}_{eq} \mathbf{F}_{BB} = \begin{bmatrix} \tilde{\mathbf{H}}_1 \\ \tilde{\mathbf{H}}_2 \\ \vdots \\ \tilde{\mathbf{H}}_K \end{bmatrix} \begin{bmatrix} \bar{\mathbf{V}}_1^{N_{RF}^r}, \dots, \bar{\mathbf{V}}_K^{N_{RF}^r} \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{H}}_1 \bar{\mathbf{V}}_k^{N_{RF}^r} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{H}}_2 \bar{\mathbf{V}}_K^{N_{RF}^r} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \tilde{\mathbf{H}}_K \bar{\mathbf{V}}_K^{N_{RF}^r} \end{bmatrix}. \tag{21}$$

The BD-baseband equivalent channel matrix \mathbf{H}_{BD} of the k -th user can be expressed as $\mathbf{H}_{BDk} = \tilde{\mathbf{H}}_k \bar{\mathbf{V}}_k^{N_{RF}^r}$. The SVD of $\tilde{\mathbf{H}}_k \bar{\mathbf{V}}_k^{N_{RF}^r}$ is defined as

$$\tilde{\mathbf{H}}_k \bar{\mathbf{V}}_k^{N_{RF}^r} = \mathbf{U}_k [\Sigma_k, \mathbf{0}] \left[\mathbf{V}_k^{N_s}, \mathbf{V}_k^{(N_{RF}^r - N_s)} \right], \tag{22}$$

where $\mathbf{V}_k^{N_s}$ represents the right singular matrix corresponding to a nonzero singular value, and $\mathbf{V}_k^{(N_{RF}^r - N_s)}$ represents the right singular matrix corresponding to a zero-singular value. Thus, we define the digital precoder as

$$\mathbf{F}_{BB} = \left[\bar{\mathbf{V}}_1^{N_{RF}^r}, \bar{\mathbf{V}}_2^{N_{RF}^r}, \dots, \bar{\mathbf{V}}_K^{N_{RF}^r} \right] \begin{bmatrix} \mathbf{V}_1^{N_s} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2^{N_s} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{V}_K^{N_s} \end{bmatrix}. \tag{23}$$

The baseband combiner of the k -th user is chosen as

$$\mathbf{W}_{RFk} = \mathbf{U}_k^{N_s}, k = 1, 2 \dots K. \tag{24}$$

Therefore, the sum spectral efficiency achieved by the proposed hybrid precoding scheme is

$$\begin{aligned} \mathbf{R} &= \sum_{k=1}^K \log_2 \left(\left| I_{N_s} + \frac{P}{KN_s} \mathbf{R}_i^{-1} \mathbf{W}_{BBk}^H \tilde{\mathbf{H}}_k \mathbf{F}_{BBk} \mathbf{F}_{BBk}^H \tilde{\mathbf{H}}_k^H \mathbf{W}_{BBk} \right| \right) \\ &= \sum_{k=1}^K \log_2 \left(\left| I_{N_s} + \frac{P}{\sigma^2 KN_s} (\mathbf{W}_{BBk}^H \mathbf{W}_{RFk}^H \mathbf{W}_{RFk} \mathbf{W}_{BBk})^{-1} \mathbf{W}_{BBk}^H \tilde{\mathbf{H}}_k \mathbf{F}_{BBk} \mathbf{F}_{BBk}^H \tilde{\mathbf{H}}_k^H \mathbf{W}_{BBk} \right| \right). \end{aligned} \quad (25)$$

The DFT orthogonal basis is used to construct the analog combining matrix \mathbf{W}_{RFk} , where $\mathbf{W}_{RFk}^H \mathbf{W}_{RFk} = \mathbf{I}$. Thus, we obtain

$$\mathbf{W}_{BBk}^H \mathbf{W}_{RFk}^H \mathbf{W}_{RFk} \mathbf{W}_{BBk} = (\mathbf{U}_k^{N_s})^H \mathbf{U}_k^{N_s} = \mathbf{I}_{N_s}. \quad (26)$$

The sum spectral efficiency of K users can finally be expressed as

$$\begin{aligned} [\mathbf{R}] &= \sum_{k=1}^K \log_2 \left(\left| I_{N_s} + \frac{P \Lambda_k (\mathbf{W}_{BBk}^H \mathbf{W}_{RFk}^H \mathbf{W}_{RFk} \mathbf{W}_{BBk})^{-1} (\boldsymbol{\Sigma}_k^{N_s})^2}{\sigma^2 KN_s} \right| \right), \\ &= \sum_{k=1}^K \log_2 \left(\left| I_{N_s} + \frac{P \Lambda_k (\boldsymbol{\Sigma}_k^{N_s})^2}{\sigma^2 KN_s} \right| \right), \end{aligned} \quad (27)$$

where Λ is a diagonal matrix, whose elements are the power loading coefficients, which can be found by water-filling on the singular values $\boldsymbol{\Sigma}_k^{N_s}$ from all users collected together, assuming a total power constraint.

4.3 Computational complexity analysis

In this subsection, the computational complexity of the proposed hybrid precoding algorithm is discussed. The complexity of the proposed algorithm comes from the assignment of analog precoders and the calculation of digital precoders. Table 1 provides the complexity comparison between the proposed hybrid precoding and some other schemes. The full-digital BD precoding scheme has the highest computational complexity. Considering the typical mmWave MIMO system with $N_t = 256, N_r = 16, K = 8, N_{RF}^t = 8, N_{RF}^r = 1$, we can observe that the complexity of the proposed BD precoding scheme is approximately 67,108,864 multiplications, which is only approximately 0.45% as complex as the full-digital BD precoding scheme.

Compared with the existing hybrid precoding algorithms based on the single-user system [23–25] and the single-antenna multi-user systems [26, 27], our work can be applied

Table 1 Computational complexity

Precoding schemes	Computational complexity
Full-digital BD	$\mathcal{O}(K^2 N_t^2 N_r)$
Spatially sparse	$\mathcal{O}((N_{RF}^t)^4 + N_t (N_{RF}^t)^3 + N_t (N_{RF}^t)^2 KN_s + N_{RF}^t N_t KN_s)$
MMSE	$\mathcal{O}(K^3 (N_{RF}^r)^4 + K^3 (N_{RF}^r)^3 N_r + K^3 (N_{RF}^r)^2 N_r N_s + K^2 N_{RF}^r N_r N_s)$
Limited feedback	$\mathcal{O}(K^3 (N_{RF}^t)^3 + KN_t N_r N_{RF}^t + KN_t N_{RF}^t N_{RF}^t)$
Hybrid BD	$\mathcal{O}(K^4 (N_{RF}^r)^3 + K (N_{RF}^r)^3 + K^2 (N_{RF}^r)^2 N_{RF}^t + K^2 N_t N_r N_{RF}^t + KN_t N_{RF}^t N_{RF}^t + K^2 N_{RF}^t N_{RF}^t N_s)$
Proposed hybrid precoding	$\mathcal{O}(K^4 (N_{RF}^r)^3 + K (N_{RF}^r)^3 + K^2 (N_{RF}^r)^2 N_t + K^2 N_t N_r N_{RF}^t + KN_t^2 N_{RF}^t + K^2 N_{RF}^t N_t N_s)$

in multi-antenna multi-user systems with low complexity and high spectral efficiency. To avoid high computational complexity caused by non-convex mixed integer optimization of joint optimization [28], the proposed precoding and combining schemes use two-stage design. Compared with the two-stage hybrid precoding algorithms in [29–32], the proposed algorithm achieves a tradeoff between system performance and complexity because the multi-user equivalent channel matrix enabled by the proposed analog precoder can directly aggregate the channel gains in massive MIMO system. In addition, compared with the traditional EGT algorithms in [30, 41, 42], a DFT basis is constructed to significantly reduce the computational complexity. Moreover, the proposed algorithm can make full use of the multi-antenna array gain to achieve a higher sum rate so that a higher spectral efficiency performance can be guaranteed. Compared with [31, 48, 49] and the conventional BD precoding schemes, the proposed algorithm has the optimal spectral efficiency and BER performance with the lowest computational complexity.

5 Simulation results and discussion

In this section, we compare the performance of the proposed algorithm with the spatially sparse hybrid precoding scheme [48], the MMSE combiner hybrid precoding scheme [49], the limited feedback hybrid precoding scheme [31], and the conventional BD precoding scheme. In addition, we also provide the simulation results of the spectral efficiency to evaluate the performance of the proposed hybrid precoding scheme in Rayleigh fading channel to verify the practicability of this algorithm in different channels, because the Rayleigh fading channel can effectively describe the wireless environment with obstacles which can scatter many signals. For mmWave channel, the clustered mmWave channel model is employed to characterize the limited scattering feature of the mmWave channel. Both the transmission and receiving antenna arrays are ULAs with antenna spacing $d = \frac{\lambda}{2}$. Since the BS usually employs directional antennas to eliminate interference and increase antenna gains [20], the AoDs are assumed to follow a uniform distribution within $[-\frac{\pi}{6}, \frac{\pi}{6}]$. Moreover, due to the random position of users, we assume that the AoAs follow a uniform distribution within $[-\pi, \pi]$, which means that omnidirectional antennas are adopted by users. For Rayleigh fading channel, variable settings are consistent with those in the mmWave channel.

5.1 Spectral efficiency

To verify the spectral efficiency performance of the proposed algorithm, Fig. 2 depicts the performance comparison of the sum spectral efficiency against the SNR in mmWave channel and Rayleigh channel, with $N_t = 128$ and $N_r = 16$. It can be seen that the proposed scheme is significantly superior to the other schemes in both channels. The EGT enabled by the RF precoder can directly aggregate the channel gains so that the spectral efficiency performance can be guaranteed. When SNR = 0 dB, the sum spectral efficiency of the proposed scheme in mmWave channel is 47.7 bits/s/Hz, and it is 47.26 bits/s/Hz in Rayleigh channel. Comparing Fig. 2a and b, we find that when the number of BS antennas is 128, the performance of the proposed scheme in the Rayleigh channel is worse than that in the mmWave channel. It is probably due to the fact that the DFT bases selection in the proposed scheme essentially captures the dominant paths of the mmWave channels.

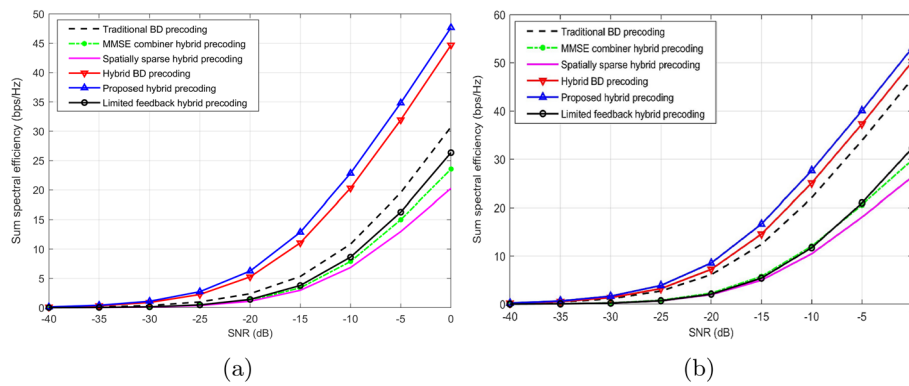


Fig. 2 Achievable spectral efficiency comparison for the MIMO system in (a) mmWave channel; (b) Rayleigh channel

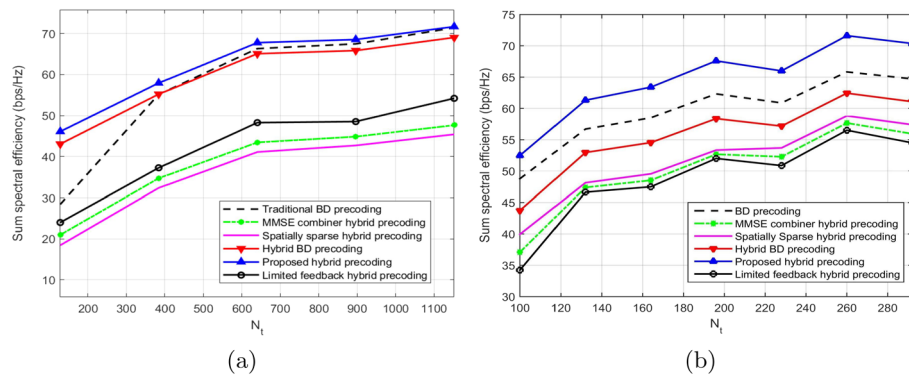


Fig. 3 Achievable spectral efficiency comparison against the number of BS antennas in (a) mmWave channel; (b) Rayleigh channel

5.2 Robustness evaluation

Figure 3 compares the sum spectral efficiency versus the number of BS antennas in both channels given $K = 8, 16$ user antennas, and $\text{SNR} = 0$ dB. The number of scattering clusters is 8, and each cluster has 10 paths so that the Rayleigh channel is complex with multiple scattered clusters and the mmWave channel is sparse. We find that the performance improves with an increase in the number of BS antennas for all hybrid structures. Moreover, because the useful information only exists in a few paths in mmWave sparse channel, it is more difficult to obtain the exact channel state information. In this case, the proposed algorithm has better spectral efficiency than the traditional BD method, so it can get better performance when used in the massive MIMO systems with different channels.

Simulation results regarding the performance for different numbers of users are provided in Fig. 4, with $N_t = 256, N_r = 16$ and $\text{SNR} = 0$ dB. It can be clearly seen that the performance in both channels improves with the increase in the number of users for all hybrid structures and the performance of the proposed scheme is optimal by using an improved low-dimensional BD precoder and combiner. When the number of users is 8, the sum spectral efficiency gap between the proposed scheme and the hybrid BD precoding scheme is 2.738 bits/s/Hz in mmWave channel and 2.136 bits/s/

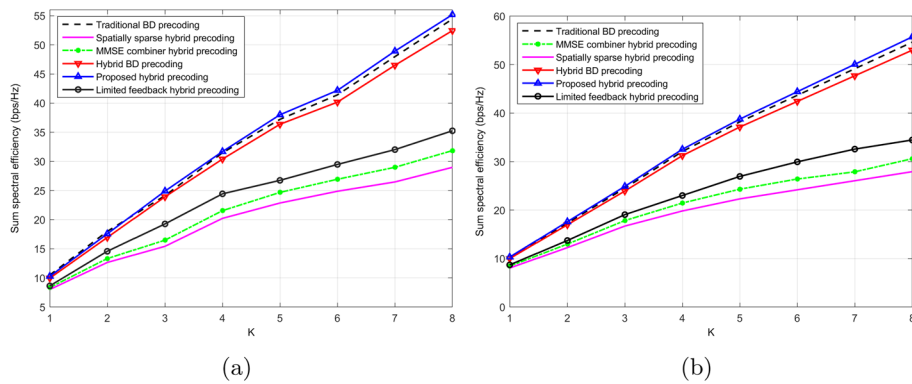


Fig. 4 Achievable spectral efficiency comparison against the number of users K in (a) mmWave channel; (b) Rayleigh channel

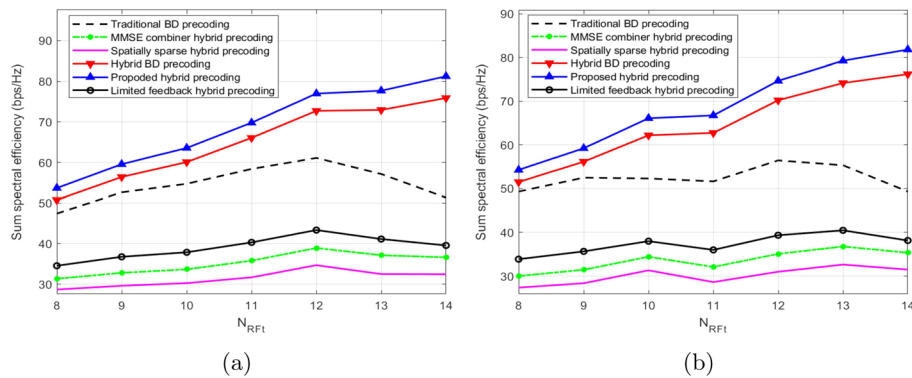


Fig. 5 Achievable spectral efficiency comparison against the number of RF chains in (a) mmWave channel; (b) Rayleigh channel

Hz in Rayleigh channel. Therefore, the proposed algorithm can significantly cancel inter-user interference in multi-user systems.

Simulation results regarding the performance for different numbers of BS RF chains are provided in Fig. 5, with $N_t = 256, N_r = 16$. It can be clearly seen that the performance improves as the number of RF chains increase for all hybrid structures and the proposed scheme is significantly superior to the other schemes. The proposed scheme can be safely recommended for implementation in systems with a large number of RF chains, because it is less vulnerable to the inter-user interference than the other schemes in this case. However, as the number of RF chains increases to 13, the performance of all the algorithms tends to be stabilized. This is because when the number of RF chains grows beyond an optimal value, inter-user interference substantially becomes more severe and the sum spectral efficiency is gradually influenced.

Simulation results regarding the performance for different numbers of data streams are provided in Fig. 6, with $N_t = 256, N_r = 16$. Figure 6 shows that when the number of data streams is 8, the spectral efficiency achieved by the proposed scheme is 130.9 bits/s/Hz in mmWave channel and 134.3 bits/s/Hz in Rayleigh channel. It can be clearly seen that the performance improves as the number of data streams for all hybrid structures increase, and that the proposed scheme is significantly superior

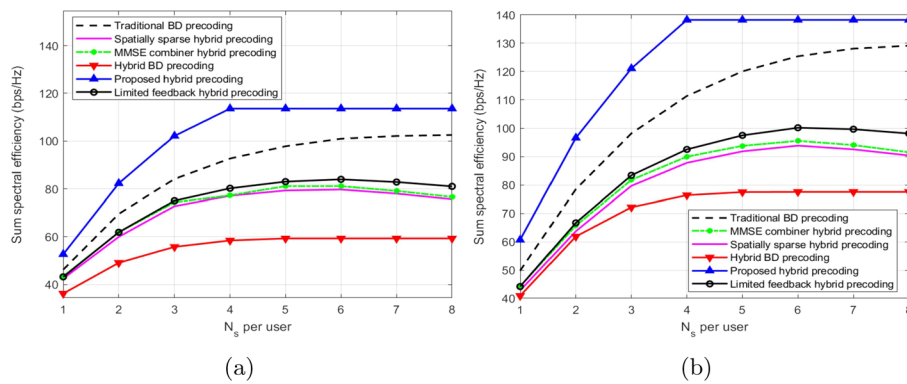


Fig. 6 Achievable spectral efficiency comparison against the number of data streams in (a) mmWave channel; (b) Rayleigh channel

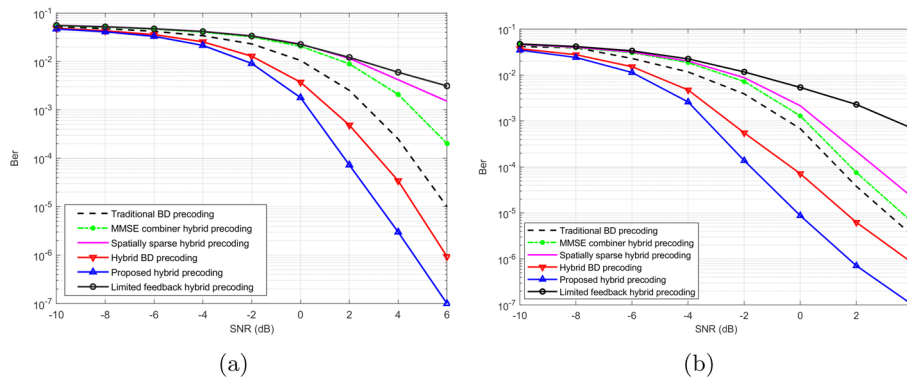


Fig. 7 BER performance comparison for the massive MIMO system in (a) mmWave channel; (b) Rayleigh channel

to the other schemes. However, as the number of data streams increases to 5, the performance of the proposed algorithm tends to be decrease in both channels. This is because the pursuit of the large array gain slightly introduces the inter-stream interference in the RF domain, which will degrade the system spectral efficiency after the baseband BD processing.

Figure 7 illustrates the bit-error-rate (BER) performance achieved by the proposed scheme for the 256×16 massive MIMO system. The proposed hybrid precoding method achieves better BER performance than other hybrid precoding methods. This is because the proposed scheme generates multiple subchannels with the equal gain transmission for each user, which improves the total BER performance. From Fig. 7, we can also find that as the SNR increases, the BER achieved by each algorithm decreases. Furthermore, by comparing the achieved robustness performances between the Rayleigh channel and mmWave channel, due to the poor scattering nature of mmWave channel, we find that the proposed scheme achieves a better robustness performance in Rayleigh channel than in mmWave channel.

6 Conclusion

In this paper, we propose a high spectral-efficiency hybrid precoding algorithm for downlink multi-user mmWave massive MIMO systems with low complexity. We extend the dimensions of the analog precoding matrix into a square matrix and find the optimal analog combiner to increase the gain of the equivalent baseband channel matrix. Then, a block diagonal (BD) precoding is performed based on the equivalent channel seen from the baseband to eliminate inter-user interference. The proposed scheme, with its lower implementation and computational complexity, achieves a capacity performance that is close to and sometimes even higher than those of conventional BD processing. The simulation results indicate that the proposed scheme can achieve a better spectral efficiency and bit-error-rate (BER) performance than other precoding schemes in both the Rayleigh fading channel and mmWave channel.

Abbreviations

5 G	The fifth generation
mmWave	Millimeter wave
MIMO	Multiple-input multiple-output
BS	Base station
PS	Phase shifter
MU	Mobile user
RF	Radio frequency
DFT	Discrete Fourier transform
CSI	Channel state information
SINR	Signal-to-noise ratio
CNN	Convolutional neural network
MU-MIMO	Multi-user MIMO
EGT	Equal gain transmission
EEMI	End-to-end mutual information
HBD	Hybrid block diagonalization
BD	Block diagonalization
SSE	Sum spectral efficiency
i.i.d.	Independent and identically distribution
AWGN	Additive white Gaussian noise
AOA	Arrival azimuth angles
DOA	Departure azimuth angles
ULA	Uniform linear array
ZF	Zero-forcing
SVD	Singular value decomposition
MMSE	Minimum mean squared error
BER	Bit error rate

Acknowledgements

The authors acknowledged the anonymous reviewers and editors for their efforts in constructive and generous feedback.

Author contributions

The algorithm proposed in this paper has been conceived by Prof. YL, B.S. QZ, and Prof. XL. Prof. YL, B.S. QZ, Dr. XL, Prof. YZ, and Prof. TQ designed the experiments. B.S. QZ, B.S. XH, and Prof. YZ performed the experiments and analyzed the results. Prof. YL, B.S. QZ, and Prof. XL wrote the paper. All authors have read and agreed to the published version of the manuscript.

Funding

This work is supported by the National Natural Science Foundation of China under grants 62161037 and 62071257 and is supported in part by the Program for Young Talents of Science and Technology in Universities of Inner Mongolia Autonomous Region under Grant NJYT-20-A11 and in part by the Natural Science Foundation of Inner Mongolia Autonomous Region under Grants 2019MS06033.

Availability of data materials

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 26 July 2021 Accepted: 13 July 2022

Published online: 26 July 2022

References

1. I.P. Roberts, J.G. Andrews, H.B. Jain et al., Millimeter-wave full duplex radios: New challenges and techniques. *IEEE Wirel. Commun.* **28**(1), 36–43 (2021). <https://doi.org/10.1109/MWC.001.2000221>
2. P. Wang, J. Fang, L. Dai et al., Joint transceiver and large intelligent surface design for massive MIMO mmWave systems. *IEEE Trans. Wireless Commun.* **20**(2), 1052–1064 (2021). <https://doi.org/10.1109/TWC.2020.3030570>
3. W. Li, W. Yang, L. Yang, H. Xiong et al., Bidirectional positioning assisted hybrid beamforming for massive MIMO systems. *IEEE Trans. Commun.* **69**(5), 3367–3378 (2021). <https://doi.org/10.1109/TCOMM.2021.3055518>
4. W. Hong, Z. Jiang, C. Yu et al., The role of millimeter-wave technologies in 5G/6G wireless communications. *IEEE J. Microwaves.* **1**(1), 101–122 (2021). <https://doi.org/10.1109/JMW.2020.3035541>
5. W. Shao, S. Zhang, H. Li et al., Angle-domain NOMA over multicell millimeter wave massive MIMO networks. *IEEE Trans. Commun.* **68**(4), 2277–2292 (2020). <https://doi.org/10.1109/TCOMM.2020.2968881>
6. W. Feng, N. Zhao, S.P. Ao et al., Joint 3D trajectory and power optimization for UAV-aided mmWave MIMO-NOMA networks. *IEEE Trans. Commun.* **69**(4), 2346–2358 (2021). <https://doi.org/10.1109/TCOMM.2020.3044599>
7. Y. Cao, N. Zhao, Y. Chen et al., Secure transmission via beamforming optimization for NOMA networks. *IEEE Wirel. Commun.* **27**(1), 193–199 (2020). <https://doi.org/10.1109/MWC.001.1900159>
8. C. Hu, X. Wang, L. Dai et al., Partially coherent compressive phase retrieval for millimeter-wave massive MIMO channel estimation. *IEEE Trans. Signal Process.* **68**, 1673–1687 (2020). <https://doi.org/10.1109/TSP.2020.2975914>
9. H. Cheng, C. Chen, C. Shen et al., The configurable hybrid precoding processor for bit-stream-based mmWave MIMO systems. *IEEE Trans. Very Large Scale Integr. Syst.* **28**(12), 2612–2622 (2020). <https://doi.org/10.1109/TVLSI.2020.3029626>
10. C.G. Tsinos, S. Chatzinotas, B. Ottersten, Hybrid analog-digital transceiver designs for multi-user MIMO mmWave cognitive radio systems. *IEEE Trans. Cognit. Commun. Netw.* **6**(1), 310–324 (2020). <https://doi.org/10.1109/TCCN.2019.2933423>
11. E. Vlachos, J. Thompson, Energy-efficiency maximization of hybrid massive MIMO precoding with random-resolution DACs via RF selection. *IEEE Trans. Wireless Commun.* **20**(2), 1093–1104 (2021). <https://doi.org/10.1109/TWC.2020.3030772>
12. A. Haqiqatnejad, F. Kayhan, B. Ottersten, Energy-efficient hybrid symbol-level precoding for large-scale mmWave multiuser MIMO systems. *IEEE Trans. Commun.* **69**(5), 3119–3134 (2021). <https://doi.org/10.1109/TCOMM.2021.3058967>
13. K. Chen, C. Qi, G. Li, Two-step codeword design for millimeter wave massive MIMO systems with quantized phase shifters. *IEEE Trans. Signal Process.* **68**, 170–180 (2020). <https://doi.org/10.1109/TSP.2019.2959250>
14. A. Arora, C.G. Tsinos, B.S.M.R. Rao et al., Hybrid transceivers design for large-scale antenna arrays using majorization-minimization algorithms. *IEEE Trans. Signal Process.* **68**, 701–714 (2020). <https://doi.org/10.1109/TSP.2019.2957611>
15. L. Dai, B. Wang, M. Peng et al., Hybrid precoding-based millimeter-wave massive MIMO-NOMA with simultaneous wireless information and power transfer. *IEEE J. Sel. Areas Commun.* **37**(1), 131–141 (2019). <https://doi.org/10.1109/JSAC.2018.2872364>
16. Y. Sun, Z. Gao, H. Wang et al., Principal component analysis-based broadband hybrid precoding for millimeter-wave massive MIMO systems. *IEEE Trans. Wireless Commun.* **19**(10), 6331–6346 (2020). <https://doi.org/10.1109/TWC.2020.3002719>
17. W. Huang, Q. Si, M. Jin, Alternating optimization based low complexity hybrid precoding in millimeter wave MIMO systems. *IEEE Commun. Lett.* **24**(3), 635–638 (2020). <https://doi.org/10.1109/LCOMM.2019.2958694>
18. N. Li, Y. Gao, K. Xu, On the optimal energy efficiency and spectral efficiency trade-off of CF massive MIMO SWIPT system. *EURASIP J. Wirel. Commun. Netw.* (2021). <https://doi.org/10.1186/s13638-021-02035-w>
19. T. Xie, L. Dai, D.W.K. Ng et al., On the power leakage problem in millimeter-wave massive MIMO with lens antenna arrays. *IEEE Trans. Signal Process.* **67**(18), 4730–4744 (2019). <https://doi.org/10.1109/TSP.2019.2926019>
20. L. Dai, B. Wang, M. Peng et al., Hybrid precoding-based millimeter-wave massive MIMO-NOMA with simultaneous wireless information and power transfer. *IEEE J. Sel. Areas Commun.* **37**(1), 131–141 (2019). <https://doi.org/10.1109/JSAC.2018.2872364>
21. C. Hu, L. Dai, T. Mir et al., Super-resolution channel estimation for mmWave massive MIMO with hybrid precoding. *IEEE Trans. Veh. Technol.* **67**(9), 8954–8958 (2018). <https://doi.org/10.1109/TVT.2018.2842724>
22. X. Xue, Y. Wang, L. Dai et al., Relay hybrid precoding design in millimeter-wave massive MIMO systems. *IEEE Trans. Signal Process.* **66**(8), 2011–2026 (2018). <https://doi.org/10.1109/TSP.2018.2799201>
23. A.W. Shaban, O. Damen, Y. Xin, et al., Statistically-aided codebook-based hybrid precoding for mmWave single-user MIMO systems, in *2020 IEEE International Conference on Communications Workshops (ICC Workshops), Dublin, Ireland*, (2020), pp. 1–6. <https://doi.org/10.1109/ICCWshops49005.2020.9145199>.

24. Z. Hong, T. Li, F. Li, Deep double-pilot-based hybrid precoding in UAV-enabled mmWave massive MIMO. *EURASIP J. Wirel. Commun. Netw.* (2020). <https://doi.org/10.1186/s13638-020-01854-7>
25. H. Nosrati, E. Aboutanios, D. Smith, Switch-based hybrid precoding in mmWave massive MIMO systems, in *27th European Signal Processing Conference (EUSIPCO)*. A Coruna, Spain **2019**, 1–5 (2019). <https://doi.org/10.23919/EUSIPCO.2019.8902688>
26. Y. Zhao, W. Xu, J. Xu et al., Analog versus hybrid precoding for multiuser massive MIMO with quantized CSI feedback. *IEEE Commun. Lett.* **24**(10), 2319–2323 (2020). <https://doi.org/10.1109/LCOMM.2020.2999646>
27. W. Ma, C. Qi, Z. Zhang et al., Sparse channel estimation and hybrid precoding using deep learning for millimeter wave massive MIMO. *IEEE Trans. Commun.* **68**(5), 2838–2849 (2020). <https://doi.org/10.1109/TCOMM.2020.2974457>
28. A.M. Elbir, A.K. Papazafeiropoulos, Hybrid precoding for multiuser millimeter wave massive MIMO systems: A deep learning approach. *IEEE Trans. Veh. Technol.* **69**(1), 552–563 (2020). <https://doi.org/10.1109/TVT.2019.2951501>
29. R. Zhang, W. Zou, Y. Wang, et al., Two-stage hybrid beamforming design for mmWave multi-user massive MIMO systems, in *IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, (2020), pp. 1–6. <https://doi.org/10.1109/WCNCW48565.2020.9124862>.
30. J. Du, Z. Wang, Y. Zhang et al., Multi-user hybrid precoding for mmWave massive MIMO systems with sub-connected structure. *EURASIP J. Wirel. Commun. Netw.* (2021). <https://doi.org/10.1186/s13638-021-02031-0>
31. A. Alkhateeb, G. Leus, R.W. Heath, Limited feedback hybrid precoding for multi-user millimeter wave systems. *IEEE Trans. Wireless Commun.* **14**(11), 6481–6494 (2015). <https://doi.org/10.1109/TWC.2015.2455980>
32. S. Gao, X. Cheng, L. Yang, Hybrid multi-user precoding for mmWave massive MIMO in frequency-selective channels, in *IEEE Wireless Communications and Networking Conference (WCNC)*, Seoul, Korea (South) 1–6 (2020). <https://doi.org/10.1109/WCNC45663.2020.9120523>
33. Y. Huang, C. Liu, Y. Song et al., DFT codebook-based hybrid precoding for multiuser mmWave massive MIMO systems. *EURASIP J. Wirel. Commun. Netw.* (2020). <https://doi.org/10.1186/s13634-020-00669-4>
34. M. Yuan, H. Wang, Y. Sun, BD-UCD-based nonlinear hybrid precoding for millimeter wave massive multiuser MIMO systems. *IEEE Commun. Lett.* **25**(3), 1010–1014 (2021). <https://doi.org/10.1109/LCOMM.2020.3034131>
35. M. Wang, F. Gao, N. Shlezinger et al., A block sparsity based estimator for mmWave massive MIMO channels with beam squint. *IEEE Trans. Signal Process.* **68**, 49–64 (2020). <https://doi.org/10.1109/TSP.2019.2956677>
36. K. Guan, B. Peng, D. He et al., Channel sounding and ray tracing for intrawagon scenario at mmWave and sub-mmWave bands. *IEEE Trans. Antennas Propag.* **69**(2), 1007–1019 (2021). <https://doi.org/10.1109/TAP.2020.3016399>
37. H. Jiang, M. Mukherjee, J. Zhou et al., Channel modeling and characteristics for 6G wireless communications. *IEEE Network* **35**(1), 296–303 (2021). <https://doi.org/10.1109/MNET.011.2000348>
38. J. Zhang, N. Garg, M. Holm et al., Design of full duplex millimeter-wave integrated access and backhaul networks. *IEEE Wirel. Commun.* **28**(1), 60–67 (2021). <https://doi.org/10.1109/MWC.001.2000199>
39. F. Bellili, F. Sotirani, W. Yu, Generalized approximate message passing for massive MIMO mmWave channel estimation with laplacian prior. *IEEE Trans. Commun.* **67**(5), 3205–3219 (2019). <https://doi.org/10.1109/TCOMM.2019.2892719>
40. J. Jin, Y. Zheng, W. Chen et al., Hybrid precoding for millimeter wave MIMO systems: A matrix factorization approach. *IEEE Trans. Wireless Commun.* **17**(5), 3327–3339 (2018). <https://doi.org/10.1109/TWC.2018.2810072>
41. L. Liang, W. Xu, X. Dong, Low-complexity hybrid precoding in massive multiuser MIMO systems. *IEEE Wirel. Commun. Lett.* **3**(6), 653–656 (2014). <https://doi.org/10.1109/lwc.2014.2363831>
42. V. Jamali, A.M. Tulino, G. Fischer et al., Scalable and energy-efficient millimeter massive MIMO architectures: Reflect-array and transmit-array antennas. *IEEE Int. Conf. Commun. (ICC)* (2019). <https://doi.org/10.1109/icc.2019.8761603>
43. M. Hanif, H. Yang, G. Boudreau et al., Practical hybrid precoding for multi-user massive MIMO systems. *IEEE Can. Conf. Electr. Comput. Eng. (CCECE)* **2016**, 1–4 (2016). <https://doi.org/10.1109/CCECE.2016.7726751>
44. S. Tsai, Equal gain transmission with antenna selection in MIMO communications. *IEEE Trans. Wireless Commun.* **10**(5), 1470–1479 (2011). <https://doi.org/10.1109/TWC.2011.030411.100410>
45. X. Yu, Y. Chu, F. Jiang et al., SVMs classification based two-side cross domain collaborative filtering by inferring intrinsic user and item features. *Knowl.-Based Syst.* **141**, 80–91 (2018). <https://doi.org/10.1016/j.knsys.2017.11.010>
46. X. Yu, F. Jiang, J. Du et al., A cross-domain collaborative filtering algorithm with expanding user and item features via the latent factor space of auxiliary domains. *Pattern Recogn.* **94**, 96–109 (2019). <https://doi.org/10.1016/j.patcog.2019.05.030>
47. F.R. Castillo-Soria, C.A. Gutierrez, A. Garcia-Barrientos et al., EQSM-based multiuser MIMO downlink transmission for correlated fading channels. *EURASIP J. Wirel. Commun. Netw.* (2020). <https://doi.org/10.1186/s13638-020-1648-1>
48. O.E. Ayach, S. Rajagopal, S. Abu-Surra et al., Spatially sparse precoding in millimeter wave MIMO systems. *IEEE Trans. Wireless Commun.* **13**(3), 1499–1513 (2014). <https://doi.org/10.1109/TWC.2014.011714.130846>
49. D.H.N. Nguyen, L.B. Le, T. Le-Ngoc et al., Hybrid MMSE precoding and combining designs for mmWave multiuser systems. *IEEE Access.* **5**, 19167–19181 (2017). <https://doi.org/10.1109/ACCESS.2017.2754979>

Publisher's Note

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