Hybridly Connected Structure for Hybrid Beamforming in mmWave Massive MIMO Systems

Didi Zhang, Yafeng Wang, Senior Member, IEEE, Xuehua Li, and Wei Xiang, Senior Member, IEEE

Abstract—In this paper, we propose a hybridly connected structure for hybrid beamforming in millimeter-wave (mmWave) massive MIMO systems, where the antenna arrays at the transmitter and receiver consist of multiple sub-arrays, each of which connects to multiple radio frequency (RF) chains, and each RF chain connects to all the antennas corresponding to the sub-array. In this structure, through successive interference cancelation, we decompose the precoding matrix optimization problem into multiple precoding sub-matrix optimization problems. Then, near-optimal hybrid digital and analog precoders are designed through factorizing the precoding sub-matrix for each sub-array. Furthermore, we compare the performance of the proposed hybridly connected structure with the existing fully and partially connected structures in terms of spectral efficiency, the required number of phase shifters, and energy efficiency. Finally, simulation results are presented to demonstrate that the spectral efficiency of the hybridly connected structure is better than that of the partially connected structure and that its spectral efficiency can approach that of the fully connected structure with the increase in the number of RF chains. Moreover, the proposed algorithm for the hybridly connected structure is capable of achieving higher energy efficiency than existing algorithms for the fully and partially connected structures.

Index Terms—MIMO, hybrid precoding, millimeter wave communications, spectral efficiency, energy efficiency.

I. INTRODUCTION

In order to meet the dramatic improvements in capacity for the upcoming fifth generation (5G) system, new emerging wireless techniques have been widely investigated such as physical layer techniques [1], [2], network densification [3], [4], and so on. Nevertheless, the problem of spectrum scarcity in current cellular systems become more and more severe. As such, the allocation of new spectral resources for commercial wireless systems is of paramount importance [5], [6].

Millimeter-wave (mmWave) communications around and above 30 GHz can achieve multigigabit data rates for indoor communications [7]. More recently, attention has been paid to the use of mmWave communications for backhaul networking between cells and mobile access within a cell [8], [9]. Meanwhile, advances in electronic components and larger unlicensed spectra motivate the wireless industry to consider mmWave as a prime candidate for outdoor cellular communications in 5G systems [10]–[12]. However, compared with current cellular systems, mmWave communications have much higher carrier frequencies. It is well known that the higher the carrier frequency is, the higher the propagation path loss is experienced in wireless communications [13]. Fortunately, the decrease in the wavelength of mmWave makes it possible to place a very large number of antennas in a much smaller region. Large antenna arrays are able to provide highly directional beamforming gains via precoding, which helps overcome the propagation path loss and increase the link reliability. Moreover, larger antenna arrays can transmit multiple streams via spatial multiplexing, which helps improve spectral efficiency. However, with the increase in the number of antennas at the transmitter, the number of radio frequency (RF) chains required by fully digital beamforming (DBF) is equal to the number of antennas, which is unrealistic in terms of cost, complexity, thermal overshoot, and implementation within a small form factor at the UE side. As such, hybrid beamforming (HBF) is considered as a promising solution to reduce the number of required RF chains [14], [15].

To date, researchers’ attention has focused on two hybrid beamforming structures, namely, the fully connected structure [16]–[22] and the partially connected structure [23]–[28], as shown in Fig. 1(a) and Fig. 1(b), respectively. In [17], the orthogonal matching pursuit (OMP) algorithm is proposed to design the analog precoder by choosing each column of the analog precoding matrix from the candidate array response vectors. Therefore, design of the OMP-based hybrid analog precoder can be viewed as a spatially sparse precoding problem, which implies that increasing spatial resolution helps improve the spectral efficiency of the system [20]. It is evident that the computational complexity is proportional to the spatial resolution. Hence, research has shifted to reduce
the computational complexity of the OMP method [19], [22]. In consideration of the hardware implementation complexity of the fully connected structure, the partially connected structure is proposed in [23]. In this structure, the analog precoding matrix is block diagonal, and each block corresponds to the precoding of a sub-array, resulting in independent precoding for each sub-array. Leveraging this property, a hybrid digital and analog precoding based successive interference cancellation (SIC) structure is proposed in [24] and [25]. Most precoder designs in the above algorithms are based on singular vectors. Given the singular vectors based precoder structure is sensitive to small changes in path length [29], [30], the directional beamforming method is proposed in [30]–[33]. However, these algorithms incur some performance loss as opposed to the algorithm based on singular vectors.

More recently, attention has shifted to energy efficiency of different HBF structures in mmWave MIMO systems [25], [28]. In [25] and [28], the spectral efficiency and energy efficiency are compared between different HBF structures. Those comparison results show that the partially connected structure outperforms the fully connected structure in energy efficiency when the number of RF chains satisfies certain conditions, and underperforms in spectral efficiency.

These comparison results motivate us to design a trade-off between spectral efficiency and energy efficiency for the hybrid digital and analog precoders. Then, we propose the hybridly connected structure, where the design degrees of freedom are much more flexible in the analog domain. Moreover, near-optimal precoder designs are presented in Section IV. Our major contributions are summarized as follows:

- We propose a hybridly connected structure for mmWave massive MIMO systems, which features a lower hardware complexity compared with the fully connected structure. On the other hand, the design degrees of freedom in the analog domain are more flexible in comparison with the partially connected structure;
- For the proposed hybridly connected structure, we propose a matrix factorization based near-optimal design for the hybrid digital and analog precoders. Meanwhile, the proposed algorithm is also applicable to the fully and partially connected structures, since those two structures are special cases of their hybridly connected counterpart;
- For the proposed hybridly connected structure, we also study the effects of various parameters such as the number of sub-arrays, RF chains on the performance of our design. Simulation results demonstrate the superiority of the proposed design.

It is worth noting that to the best of the authors’ knowledge, our work is the first to consider the design of HBF using the hybridly connected structure in mmWave MIMO systems.

The remainder of the paper is organized as follows. The system model of the mmWave massive MIMO system is presented in Section II. The hybrid precoding matrix in the hybridly connected structure is elaborated in Section III. The design of constraint hybrid digital and analog precoders is presented in Section IV. Simulation results on the spectral efficiency and energy efficiency of various HBF structures and the sensitivity of the proposed algorithm to channel estimation errors are presented in Section V. Finally, concluding remarks are drawn in Section VI.

**Notations:** Upper-case bold and lower-case bold letters represent matrices and vectors, respectively. $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^\dagger$, and $(\cdot)^{-1}$ refer to the transpose, conjugate transpose, pseudoinverse, inversion of a matrix, respectively. $|\cdot|$ is taken to mean the determinant of a matrix, and $\|\cdot\|_F$ indicates the Frobenius norm of a matrix.

### II. System Model

In this section, three structures for hybrid precoding in mmWave MIMO systems are illustrated in Fig. 1. The fully connected structure is shown in Fig. 1(a), where the transmitter is equipped with $N_R$ RF chains and $N_t$ antennas. The partially connected structure is presented in Fig. 1(b), where the transmitter is equipped with $N_R$ RF chains and $N_R M$ antennas, and each sub-array is connected to one RF chain. The hybridly connected structure is illustrated in Fig. 1(c), where the transmitter is equipped with $D$ sub-arrays, $S D$ RF chains and $D N$ antennas, and each sub-array is connected to $S$ RF chains. Moreover, the number of RF chains at the transmitter satisfies $N_t \leq SD \leq N_t$. In order to assess the performance of these hybrid precoding structures, it is assumed that all the transmitters have the same number of antennas (i.e., $N_t = N_R M = DN$), and transmit $N_t$ independent data streams to the receiver with $N_r$ receive antennas. To date,
research efforts have focused on the fully and partially connected structures, whereas the hybridly connected structure has not been studied in the literature. This paper aims to investigate hybridly connected structure.

In the hybridly connected structure, the digital precoder \( F_B \) first assigns \( N_t \) independent data streams to different RF chains, then the analog precoder \( F_R \) maps the processed signals to the corresponding sub-array antennas. Moreover, \( F_B \) is an \( SD \times N_t \) matrix, while \( F_R \) is of dimension \( N_t \times SD \). Therefore, the received signal vector \( y = [y_1, y_2, \cdots, y_{N_t}]^T \) at the receiver can be written as

\[
y = \sqrt{\gamma} H F_R F_B s + n = \sqrt{\rho} H F s + n, \quad (1)
\]

where \( \gamma \) indicates the average received power, \( H \in \mathbb{C}^{N_t \times N_t} \) stands for the channel matrix, \( s \) is the \( N_t \times 1 \) data stream vectors such that \( E [s s^H] = \frac{1}{N_t} I_{N_t} \), \( n \) is the noise vector, that follows a complex Gaussian distribution, i.e., \( n \in \mathcal{C} \mathcal{N}(0, \sigma_n^2) \), and the hybrid precoders \( F_R F_B \) should satisfy \( \| F_R F_B \|_F^2 \leq N_s \) to meet the total transmit power constraint. In this paper, we assume that the channel state information (CSI) is perfectly known at both the transmitter and receiver. In practical systems, CSI at the receiver can be obtained via channel estimation [34]–[36], and is timely shared at the transmitter via an effective feedback strategy [37]–[39].

Considering the mmWave channels have a sparse scattering structure [40]–[42], in this paper we adopt a narrow-band channel model, i.e., the extended Saleh-Valenzuela model [43]. The channel matrix \( H \) can be expressed as

\[
H = \sqrt{\frac{N_t}{L}} \sum_{l=1}^{L} a_l(\phi_l,t) a_l(\phi_l,r)^H, \quad (2)
\]

where \( \gamma = \sqrt{\frac{N_t^2}{L}} \) is a normalization factor that satisfies \( \| H \|_F^2 = N_t N_r \), \( L \) represents the number of scattering paths, and \( a_l \) stands for the complex gain of the \( l \)-th path. We assume that \( a_l \) follows the complex Gaussian distribution \( \mathcal{C} \mathcal{N}(0, 1) \).

Finally, \( a_l(\phi_l,t) \) and \( a_l(\phi_l,r) \) are the antenna array response vectors at the transmitter and receiver, where \( \phi_l,t \) and \( \phi_l,r \) are the \( l \)-th path’s azimuth angle of arrival (AoA) and angle of departure (AoD), respectively. For the uniform linear array (ULA) the elevation AoA and AoD are \( \phi_l \), which can be ignored. It is assumed that the antennas of ULA are deployed along \( y \)-axis at the transmitter and receiver, the array steering vectors \( a_l(\phi_l,t) \) and \( a_l(\phi_l,r) \) are given as [44]

\[
a_l(\phi_l,t) = \frac{1}{\sqrt{N_t}} [1, e^{j k d_l \sin(\phi_l,t)}, \ldots, e^{j k d_t (N_t - 1) \sin(\phi_l,t)}]^T, \quad (3)
\]

\[
a_l(\phi_l,r) = \frac{1}{\sqrt{N_t}} [1, e^{j k d_r \sin(\phi_l,r)}, \ldots, e^{j k d_r (N_t - 1) \sin(\phi_l,r)}]^T, \quad (4)
\]

where \( k = \frac{2 \pi}{\lambda} \), \( \lambda \) is the signal wavelength, and \( d_t \) and \( d_r \) indicate the spacing of two adjacent ULA elements at the transmitter and receiver, respectively.

### III. Hybrid Precoding Matrix for the Hybridly Connected Structure

This section elaborates on the structure of the hybrid precoding matrix \( F \) in the hybridly connected structure as shown in Fig. 1(c). It is known that the size of the hybrid precoding matrix \( F \) depends on the number of transmit data streams \( N_s \) and the number of antennas \( N_t \) at the transmitter. When both \( N_s \) and \( N_t \) are a priori known, the dimension of the hybrid precoding matrix \( F \) is fixed, i.e., \( F \in \mathbb{C}^{N_t \times N_t} \). For the hybridly connected structure, the analog precoding matrix \( F_R \) is a block diagonal matrix, which means hybrid analog precoding is independent for each sub-array. Moreover, the digital precoding matrix \( F_B \) is an \( SD \times N_t \) matrix, whose numbers of rows and columns correspond to the numbers of RF chains and streams, respectively. Therefore, the mapping relationship between the streams and the \( i \)-th sub-array depends on the \( i \)-th \( S \times N_t \)-sub-matrix of the digital precoding matrix. It can be inferred from \( F = F_B F_R \) that the precoding for the \( i \)-th sub-array depends on the \( i \)-th \( N_s \times N_t \)-sub-matrix of the hybrid precoding matrix \( F \). Therefore, \( F \) can be expressed as

\[
F = \begin{bmatrix}
f_{11} & f_{12} & \cdots & f_{1N_s} \\
f_{21} & f_{22} & \cdots & f_{2N_s} \\
\vdots & \vdots & \ddots & \vdots \\
f_{D1} & f_{D2} & \cdots & f_{DN_s}
\end{bmatrix},
\]

where \( f_{ij} \in \mathbb{C}^{N_t \times 1} \) indicates the precoding for the \( j \)-th stream transmitted by the \( i \)-th sub-array. If \( f_{ij} = 0 \), this implies the \( j \)-th stream is not transmitted by the \( i \)-th sub-array. Thus, the precoding for the \( i \)-th sub-array can be expressed as \( F_{i,\text{sub}} = [f_{i1}, f_{i2}, \cdots, f_{iN_s}] \), and the precoding for the \( j \)-th stream can be represented by \( F_{(i,j)} = [f_{i1}^T, f_{i2}^T, \cdots, f_{iN_s}^T]^T \).

Based on the above discussion, the structure of the hybrid precoding matrix \( F \) in the hybridly connected structure depends on three factors, namely the number of RF chains of each sub-array, the number of sub-arrays, and the mapping method of the streams in such a structure. In order to simplify the subsequent expression, we put forward two assumptions for the allocation of the streams to simplify the structure of \( F \):

**Assumption 1:** The number of data streams transmitted by each sub-array is equal to the number of RF chains connected by each sub-array; and

**Assumption 2:** The data streams are allocated to adjacent sub-arrays on a priority basis, which means each data stream is transmitted by adjacent sub-arrays.

It is well known that a digital precoder can adjust the number of data streams transmitted by each sub-array according to the actual requirements of the system. Denote by \( N_s \) the number of data streams transmitted by the \( i \)-th sub-array. Next we use two specific cases to elaborate on the rationality of Assumption 1.

**Case 1** \((S = N_s)\): In this case, the actually required number of RF chains for each sub-array is equal to the number of data streams allocated by the digital precoder. This is because when the number of data streams transmitted by each sub-array is constant, increasing the number of RF chains of each sub-array has little effect on spectral efficiency, but will reduce the energy efficiency of the system.

**Case 2** \((S < N_s, DS \geq N_s)\): In this case, when the number of data streams transmitted by each sub-array is more than the number of RF chains connected by each sub-array, i.e., \( N_s < N_t \), the digital precoder for each sub-array is a column
full rank matrix. Moreover, the analog precoder \( F_R \) is a block diagonal matrix, which satisfies constant modulus constraints. Unfortunately, these non-convex constraints render the precoding optimal problem computationally intractable. Assumption 1 helps decompose the complex precoding optimal problem into multiple precoding optimal sub-problems, which will be illustrated in Section V.

For Assumption 2, it is worth pointing out that the proposed algorithm is equally applicable to the case where each data stream is transmitted by non-adjacent sub-arrays. Note that the allocation of the data streams based upon the above two assumptions is not unique. However, this is not the main objective of this paper, of which the focus is to design the optimal hybrid digital and analog precoders given a fixed allocation scheme of the data streams.

According to the relationship among the number of streams \( N_s \), the number of sub-arrays \( D \), and the number of RF chains \( S \) connected by each sub-array, the structure of the hybrid digital precoder \( F_B \) can be classified into two cases. In the first case, the product of \( D \) and \( S \) is equal to \( N_s \), i.e., \( N_s = DS \). In this case, the hybrid digital precoder \( F_B \) is a block diagonal matrix. In the second case, assuming \( D \) is fixed, the product of \( D \) and \( S \) satisfies \( DS > N_s \). In this case, the hybrid digital precoder \( F_B \) is an irregular matrix. Then, since \( F = F_Q F_B \), the structure of matrix \( F \) is similar to matrix \( F_B \) in column, which can also be divided into two cases. When \( N_s = DS \), \( F \) is a block diagonal matrix. When \( DS > N_s \), \( F \) is an irregular matrix. To further illustrate the structure of the hybrid precoder \( F \) in the hybridly connected structure, we assume \( N_s = 8 \), and a specific example is given as follows.

1) Case 1 \((N_s = DS)\): when \( N_s = DS = 8 \), we have \( D \in \{1, 2, 4, 8\} \). Note that when \( D = 1 \) and \( S = 8 \), the hybridly connected structure is the same as the fully connected structure. Moreover, when \( D = 8 \) and \( S = 1 \), the hybridly connected structure is identical to the partially connected structure. That is, the fully and partially connected structures are special cases of the hybridly connected structure. Therefore, we focus on the situation of \( D = 2 \) and \( D = 4 \).

As discussed above, we know that when \( D = 2 \) and \( D = 4 \), the hybrid digital and analog precoders \( F_B \) and \( F_R \) are both block diagonal matrices. For example, if \( D = 4 \), then we have \( S = 2 \). That is, each sub-array transmits two independent data streams. The hybrid digital precoding matrix can be written as

\[
F_B = \begin{bmatrix}
F_{B,11} & F_{B,12} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & F_{B,23} & F_{B,24} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & F_{B,35} & F_{B,36} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & F_{B,47} & F_{B,48}
\end{bmatrix},
\]

where \( f_{B,ij} \in \mathbb{C}^{2 \times 1} \). Since \( F = F_R F_B \), we know that \( F \) is also a block diagonal matrix as shown in Fig. 2(a), where \( f_{ij} = f_{R,ij} f_{B,ij} \) and \( f_{ij} \in \mathbb{C}^{N \times 1} \).

2) Case 2 \((N_s < DS)\): For this case, we assume \( D = 4 \) and \( S \in \{3, 4, 5, 6, 7, 8\} \). In this situation, it is known from the above discussions that the analog precoder \( F_R \) is a block diagonal matrix, i.e., \( F_R = \text{diag}(F_{R,1}, F_{R,2}, F_{R,3}, F_{R,4}) \), and the digital precoder \( F_B \) is an irregular matrix. It can be inferred from \( F = F_R F_B \) that \( F \) is an irregular matrix. For example, if \( S = 6 \), meaning each sub-array transmits six independent data streams, one example of the digital precoding matrix \( F_B \) is given as

\[
F_B = \begin{bmatrix}
F_{B,11} & F_{B,12} & F_{B,13} & F_{B,14} & F_{B,15} & F_{B,16} & 0 & 0 \\
F_{B,21} & F_{B,22} & F_{B,23} & F_{B,24} & F_{B,25} & F_{B,26} & 0 & 0 \\
0 & 0 & F_{B,33} & F_{B,34} & F_{B,35} & F_{B,36} & F_{B,37} & F_{B,38} \\
0 & 0 & F_{B,43} & F_{B,44} & F_{B,45} & F_{B,46} & F_{B,47} & F_{B,48}
\end{bmatrix},
\]

where \( f_{B,ij} \in \mathbb{C}^{6 \times 1} \). According to \( F = F_R F_B \), the structure of \( F \) is shown in Fig. 2(c). Moreover, when \( S = 4 \) and \( S = 8 \), the structure of are shown in Fig. 2(b) and Fig. 2(d), respectively.

**IV. DIGITAL AND ANALOG PRECODERS DESIGN FOR HYBRIDLY CONNECTED STRUCTURE**

In this section, we describe the design process of the digital and analog precoding matrices \( F_B \) and \( F_R \) for the hybridly connected structure as shown in Fig. 1(c). Firstly, according to the structure of the hybrid precoding matrix, we consider obtaining unconstrained hybrid precoding matrix using SIC. Then, we design the hybrid digital and analog precoding matrices for each sub-array according to the corresponding factorization of the hybrid precoding sub-matrix.

**A. SIC-Based Unconstrained Hybrid Precoding Matrix Design**

In this subsection, we design the unconstrained hybrid precoding matrix \( F \) through maximizing the total achievable
rate of the mmWave MIMO system. As discussed in Section II, the total achievable rate of the hybridly connected mmWave MIMO system as shown in Fig. 1(c) can be expressed as

\[ C = \log_2 \left( \left| I_N + \frac{\rho}{N_{3} \sigma^2} H F H^H \right| \right), \]  

(8)

where the hybrid precoding matrix \( F \) is attainable through solving

\[ F_{\text{opt}} = \arg \max_{F} \log_2 \left( \left| I_N + \frac{\rho}{N_{3} \sigma^2} H F H^H \right| \right), \]  

(9)

As discussed in Section III, \( F \) is a block diagonal matrix if \( N_3 = DS \), while \( F \) is an irregular matrix if \( N_3 < DS \). Meanwhile, the inequality of \( \| F \|_F \leq N_0 \) should be satisfied to meet the total transmit power constraint. Unfortunately, these non-convex constraints make the optimal problem (9) computationally intractable to solve. Since each column of \( F \) represents the precoding vector of one stream, we consider decomposing the optimization problem (9) into multiple sub-rate optimization problems. Moreover, considering that different streams are transmitted by the same multiple sub-arrays, we decompose \( F \) into multiple sub-matrices. For example, when \( D = 4 \) and \( S = 6 \), the structure \( F \) is shown in Fig. 2(c). As can be observed from the figure, the first and second streams are only transmitted by the first and second sub-arrays, while the third to sixth streams are transmitted by all the sub-arrays. The seventh and eighth streams are only transmitted by the third and fourth sub-arrays. Therefore, matrix \( F \) can be partitioned into three sub-matrices as shown in Fig. 2(c). Moreover, the partition results of the hybrid precoding matrix for \( S = (2, 4, 8) \) are presented in Fig. 2(a), (b) and (d).

According to the above discussions, we divide the hybrid precoding matrix into \( K \) sub-matrices \( F = \{ F_1, F_2, \ldots, F_K \} \) and \( F_k \in C^{N_{k} \times k} \), where \( N_k \) is the number of streams precoded by the \( k \)th sub-matrix. Then \( F \) can be further expressed as \( F = [ \hat{F}_{k-1} F_K ] \), where \( \hat{F}_{k-1} \) denotes the first \( K-1 \) sub-matrices of \( F \). Furthermore, the total achievable rate \( C \) in (8) can be rewritten as

\[ C = \log_2 \left( \left| I_N + \frac{\rho}{N_{3} \sigma^2} H F K H^H \right| \right) = \log_2 \left( \left| I_N + \frac{\rho}{N_{3} \sigma^2} H [ \hat{F}_{k-1} F_K ] [ \hat{F}_{k-1} F_K ]^H H^H \right| \right) \]

\[ = \log_2 \left( \left| I_N + \frac{\rho}{N_{3} \sigma^2} [H F K K^H + H \hat{F}_{k-1} \hat{F}_{k-1} H^H] \right| \right) \]

\[ \geq \sum_{k=1}^{K} \log_2 \left( \left| I_N + \frac{\rho}{N_{3} \sigma^2} R_{k-1}^{-1} H F K F K^H H^H \right| \right) \]

\[ \geq \sum_{k=1}^{K} \log_2 \left( \left| I_N + \frac{\rho}{N_{3} \sigma^2} R_{k-1}^{-1} F K F K^H \right| \right), \]  

(10)

where \( R_0 = I_N \), and \( R_{k-1} = \frac{\rho}{N_{3} \sigma^2} \hat{F}_{k-1} \hat{F}_{k-1} H^H \).

Step (a) is obtained due to the fact that \( \| X Y \|_F = \| X \|_F \| Y \|_F \) and let \( X = R_{K-1} \) and \( Y = I_N + \frac{\rho}{N_{3} \sigma^2} R_{K-1}^{-1} H F K F K^H H^H \). Note that the form of \( \log_2 (|R_{K-1}|) \) is similar to (8), and that we can use the similar method in (8) to decompose it. Therefore, step (b) is the result of \( K-1 \) iterations. Step (c) is obtained due to the fact that \( |I + X Y| = |I + Y X| \), where \( X = R_{k-1} H F K \) and \( Y = F K H^H \).

As can be observed from (10), the total achievable rate is the sum of the sub-rates of all the streams, which implies that the precoding matrix optimization problem (9) can be decomposed into a series of precoding sub-matrix optimization problems. Motivated by the idea of SIC, we can first optimize the achievable sub-rate of the first \( s_1 \) streams and update matrix \( R_2 \). This means that the phase shifters connected by the RF chains corresponding to the first \( s_1 \) streams are computed and the effect of the first \( s_1 \) streams on the other streams can be nullified by \( R_2 \). Then, similar operations can be performed to optimize the achievable sub-rate of the next \( s_2 \) streams and update matrix \( R_3 \). Repeat this process until the last \( s_K \) streams are considered. It follows from (10) that the precoding sub-matrix \( F_k \) can be obtained by solving

\[ F_{k}^{\text{opt}} = \arg \max_{F_k} \log_2 \left( \left| I_N + \frac{\rho}{N_{3} \sigma^2} R_{k-1}^{-1} F_k H \right| \right), \]  

(11)

where \( Q_k = H F_k R_k^{-1} H^H \) is a \( N_1 \times N_1 \) Hermitian matrix. We assume that the \( s_k \) streams precoded by \( F_k \) are transmitted by the \( n_k \)th to \( m_k \)th sub-arrays. Then, the hybrid precoding sub-matrix \( F_k \) can be written as \( F_k = \tilde{F}_k \left[ 0 \tilde{F}_k^H \right] \), where \( \tilde{F}_k \in C^{(m_k - n_k + 1)N \times s_k} \). Therefore, the optimization problem of (11) can be transformed into

\[ \tilde{F}_{k}^{\text{opt}} = \arg \max_{\tilde{F}_k} \log_2 \left( \left| I_N + \frac{\rho}{N_{3} \sigma^2} \tilde{F}_k \tilde{Q}_{k-1} \tilde{F}_k^H \right| \right), \]  

(12)

where \( \tilde{Q}_{k-1} = (m_k - n_k + 1)N \times (m_k - n_k + 1)N \) Hermitian matrix formed as a sub-matrix of matrix \( Q_{k-1} \) by taking the \((n_k - 1)N + 1)\)th row and column to the \((m_k N)\)th row and column of \( Q_{k-1} \). Let us define the singular value decomposition (SVD) of matrix \( Q_k \) as \( Q_k = U \Sigma U^H \), where \( U \) is a unitary matrix, \( \Sigma \) is a diagonal matrix of the singular values arranged in decreasing order. Then, matrices \( \Sigma \) and \( U \) can be partitioned as

\[ \Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix}, \quad U = \begin{bmatrix} U_1 & U_2 \end{bmatrix}, \]  

(13)

where \( \Sigma_1 \) is an \( s_k \times s_k \) diagonal matrix, and \( U_1 \) is an \((m_k - n_k + 1)N \times s_k \) matrix. Therefore, the optimal unconstrained precoding matrix of (12) can be expressed as

\[ \tilde{F}_{k}^{\text{opt}} = U_1. \]  

(14)

It is known from (14) that the optimal method described above for deriving the \( k \)th pre-matrix sub-matrix is reusable to optimize the \((k + 1)\)th pre-matrix sub-matrix. Then, we can obtain \( \tilde{F}_{k}^{\text{opt}} \) through \( K \) iterations.

As mentioned above, we need to obtain the sub-matrix \( Q_k \) by updating \( Q_k = H F_k R_k^{-1} H \) in each iteration, which is very complicated. This is because that this process involves the multiplication and inversion of a large scale matrix. Next, we focus on simplifying the calculation of \( Q_k \). Firstly, \( R_k \) can be
Thus, it is assumed that the constrained precoding matrix $G$ can be written as $G = HZ^\top F_k F_k^H H^H$, where $Z$ is a diagonalized matrix, it follows from the Sherman-Morrison formula [45, eq. (2.1.4)] that

$$\begin{align*}
(Z + XY^T)^{-1} &= Z^{-1} - Z^{-1}X(I + Y^T Z^{-1}X)^{-1}Y^T Z^{-1},
\end{align*}
$$

Therefore, $Q_k$ can be expressed as

$$Q_k = H^T R_k^{-1} H = Q_{k-1} - \frac{\rho}{N_\sigma^2} F_k^H H^T R_{k-1} H F_k \times \left(I + \frac{\rho}{N_\sigma^2} F_k^H H^T R_{k-1} H F_k \right)^{-1} F_k^H H^T R_{k-1}^{-1}.
$$

Since $F_k^H Q_{k-1} F_k = \Sigma_1$, $Q_k$ can be written as

$$Q_k = Q_{k-1} - \frac{\rho}{N_\sigma^2} F_k^H Q_{k-1} F_k \left(I + \frac{\rho}{N_\sigma^2} \Sigma_1 \right)^{-1} F_k^H Q_{k-1}.
$$

Therefore, $Q_k$ can be obtained by solving the following optimization problem:

$$\begin{align*}
\tilde{G}_k &= \arg \min_{\tilde{G}_k} \| \tilde{F}_k - \tilde{G}_k \|_F^2.
\end{align*}
$$

Note that this result can be applied to sub-matrix $G_{k+1}$. Therefore, $G$ can be obtained by solving the following optimization problem:

$$\begin{align*}
G^\text{opt} &= \arg \min_{G} \| F - G \|_F^2, \\
\text{s.t.} \ & \| G \|_F^2 \leq N_s, \\
F_R &\in \mathcal{S},
\end{align*}
$$

where $\mathcal{S}$ is a $DN \times DS$ block diagonal matrix set with constant modulus entries. As aforementioned, we then decompose the total precoding matrix optimization problem into $D$ precoding sub-matrix optimization problems for each sub-array. Thus, the precoding matrix optimization problem for the $i$th sub-array can be expressed as

$$\begin{align*}
G^\text{opt}_{i,\text{sub}} &= \arg \min_{G_{i,\text{sub}}} \| F_{i,\text{sub}} - G_{i,\text{sub}} \|_F^2,
\end{align*}
$$

where $G_{i,\text{sub}}$ refers to the constraint precoding sub-matrix of the $i$th sub-array. Moreover, the structure of $G_{i,\text{sub}}$ is similar to $F_{i,\text{sub}}$, i.e., $G_{i,\text{sub}} = [0, G_i, 0]$. Therefore, the precoding sub-matrix optimization problem (19) can be further expressed as

$$\begin{align*}
\tilde{G}_i^\text{opt} &= \arg \min_{\tilde{G}_i} \| \tilde{F}_i - \tilde{G}_i \|_F^2.
\end{align*}
$$

Algorithm 1: SIC-Based Optimal Hybrid Precoding Algorithm

**Require:** $H$, $N$, and $N_s$

**Initialization:** $R_0 = I_{N_\tau}$, $Q_0 = H^H H$

1: for $1 \leq k \leq K$ do
2: $\tilde{Q}_{k-1} = Q_{k-1} \cdot H(N_\tau - 1) + m_k N, (N_\tau - 1) + m_k N)$
3: $\tilde{Q}_k = \tilde{Q}_{k-1} - U \Sigma U$, $U = \{U_1 U_2\}$, $U_1 \in \mathbb{C}((m_k - 1) N \times k)$
4: $\tilde{F}_k = U_1$
5: $\tilde{F}_k = \tilde{Q}_{k-1} + m_k N + 1 : m_k N, 1 : s_k = F_{k-1}^\text{opt}$
6: $\tilde{Q}_k = \tilde{Q}_{k-1} - \frac{\rho}{N_\sigma^2} \tilde{F}_k \tilde{Q}_{k-1} \left(I + \frac{\rho}{N_\sigma^2} \Sigma_1 \right)^{-1} \tilde{F}_k H^T Q_{k-1}$
7: end for

**Return:** $F$
where \( \mathbf{B} = (\mathbf{F}^H \mathbf{F}_i)^{\frac{1}{2}}, \mathbf{W} = (\mathbf{G}^H \mathbf{G}_i)^{\frac{1}{2}}, \mathbf{P} \) and \( \mathbf{M} \) are \( N \times S \) matrices, both consisting of orthonormal column vectors, i.e., \( \mathbf{P}^H \mathbf{P} = \mathbf{I}_S \) and \( \mathbf{M}^H \mathbf{M} = \mathbf{I}_S \). Defining \( \mathbf{B} = [\mathbf{B}_1, \mathbf{B}_2] \) and \( \mathbf{W} = [\mathbf{W}_1, \mathbf{W}_2] \), we have

\[
\begin{align*}
\mathbf{F}_i &= \mathbf{B}_1, \quad \mathbf{F}_2 = \mathbf{B}_2 \\
\mathbf{G}_i &= \mathbf{M} \mathbf{W}_1, \quad \mathbf{G}_2 = \mathbf{M} \mathbf{W}_2.
\end{align*}
\]

Note that for massive MIMO systems, the optimal precoding matrix \( \mathbf{F} \) satisfies \( \mathbf{F}^H \mathbf{F} \approx \mathbf{I}_N \), and each column of \( \mathbf{F} \) is a unit vector. Thus, \( \mathbf{B} \) and \( \mathbf{W} \) can be expressed as

\[
\begin{align*}
\mathbf{B} &= \left( \begin{bmatrix} \mathbf{F}^H \mathbf{F}_1 & \mathbf{F}^H \mathbf{F}_2 \end{bmatrix} \right)^{\frac{1}{2}} \\
\mathbf{W} &= \left( \begin{bmatrix} \mathbf{G}^H \mathbf{G}_1 & \mathbf{G}^H \mathbf{G}_2 \end{bmatrix} \right)^{\frac{1}{2}}
\end{align*}
\]

Moreover, the objective function \( \|\mathbf{F}_i - \mathbf{G}_i\|_F^2 \) in (25) can be transformed to

\[
\begin{align*}
\|\mathbf{F}_i - \mathbf{G}_i\|_F^2 &= \text{tr} \left( (\mathbf{F}_i - \mathbf{G}_i)^H (\mathbf{F}_i - \mathbf{G}_i) \right) \\
&= \|\mathbf{F}_i\|_F^2 + \|\mathbf{G}_i\|_F^2 - 2 \text{tr} \left( \mathbf{F}_i^H \mathbf{G}_i \right) \\
&= \left( \begin{bmatrix} \mathbf{F}_{i1}^H & \mathbf{F}_{i2}^H \end{bmatrix} \right)^{\frac{1}{2}} \left[ 1 \sqrt{\frac{\beta_1}{\gamma_1}} \mathbf{I} \_1 \sqrt{\frac{\beta_2}{\gamma_2}} \mathbf{I} \_2 \right] \left( \begin{bmatrix} \mathbf{F}_{i1} & \mathbf{F}_{i2} \end{bmatrix} \right)^{\frac{1}{2}} \\
&\approx 2 \left( \frac{\beta_1}{\gamma_1} + \frac{\beta_2}{\gamma_2} \right) \text{tr} \left( \mathbf{B}_1^H \mathbf{P}^H \mathbf{M} \mathbf{W}_1 + \mathbf{B}_2^H \mathbf{P}^H \mathbf{M} \mathbf{W}_2 \right).
\end{align*}
\]

It is observed that (28) can be minimized by maximizing \( \text{tr} (\mathbf{B}_1^H \mathbf{P}^H \mathbf{M} \mathbf{W}_1) \) and \( \text{tr} (\mathbf{B}_2^H \mathbf{P}^H \mathbf{M} \mathbf{W}_2) \). Define \( \mathbf{P} = [\mathbf{P}_1, \mathbf{P}_2] \) (\( \mathbf{M} = [\mathbf{M}_1, \mathbf{M}_2] \)), where \( \mathbf{P}_1 \in \mathbb{C}^{N \times \beta_1} (\mathbf{M}_1 \in \mathbb{C}^{N \times \beta_1}) \) and \( \mathbf{P}_2 \in \mathbb{C}^{N \times \beta_2} (\mathbf{M}_2 \in \mathbb{C}^{N \times \beta_2}) \). Note that \( \mathbf{B}_1 \approx \mathbf{W}_1 \approx \left[ \begin{bmatrix} \mathbf{I} \_1 \mathbf{B}_1 \end{bmatrix} \right]^T \) and \( \mathbf{B}_2 \approx \mathbf{W}_2 \approx \left[ 0 \right]_{\sqrt{\frac{\beta_1}{\gamma_2}}}^T \). Therefore, (29) can be further rewritten as

\[
\|\mathbf{F}_i - \mathbf{G}_i\|_F^2 \approx 2 \left( \frac{\beta_1}{\gamma_1} + \frac{\beta_2}{\gamma_2} \right) \text{tr} \left( \mathbf{B}_1^H \mathbf{P}^H \mathbf{M} \mathbf{W}_1 + \mathbf{B}_2^H \mathbf{P}^H \mathbf{M} \mathbf{W}_2 \right).
\]

This is equivalent to maximizing \( \text{tr} (\mathbf{P}_1^H \mathbf{M} \mathbf{W}_1) \) and \( \text{tr} (\mathbf{P}_2^H \mathbf{M} \mathbf{W}_2) \), which is equivalent to maximizing \( \text{tr} (\mathbf{P}_1^H \mathbf{M}) \). Note that \( \mathbf{P}^H \mathbf{P} = \mathbf{I}_S \) and \( \mathbf{M}^H \mathbf{M} = \mathbf{I}_S \). Moreover, each element of \( \mathbf{M} \) has the same amplitude of \( 1/\sqrt{N} \). Therefore, the optimal analog precoding matrix \( \mathbf{F}_{R,i} \) for the \( i \)th sub-array can be expressed as

\[
\mathbf{F}_{R,i} = \frac{1}{\sqrt{N}} e^{j \angle (p)}.
\]

When \( \mathbf{F}_{R,i} \) is fixed, the digital precoding sub-matrix for the \( i \)th sub-array \( \mathbf{F}_{B,i} \) can be written as

\[
\mathbf{F}_{B,i} = \mathbf{F}_{R,i}^H \mathbf{F}_i.
\]

As can be seen from (30) and (31), \( \mathbf{F}_{R,i} \approx \mathbf{P} \) and \( \mathbf{F}_{B,i} \approx \mathbf{B} \) when \( N = \infty \). Since \( \gamma_1 \ll N \) and \( \gamma_2 \ll N \), we can conclude that the digital precoder for the \( i \)th sub-array is approximately an identity matrix, i.e., \( \mathbf{F}_{B,i} \approx \mathbf{I}_S \), when the number of antennas goes to infinity. To satisfy the power constraint, we normalize \( \mathbf{F}_{B,i} \) by a factor of \( \frac{1}{\|\mathbf{F}_{B,i}\|_F} \) to yield

\[
\mathbf{F}_{B,i}^\text{norm} = \frac{\mathbf{F}_{B,i}}{\|\mathbf{F}_{R,i} \mathbf{F}_{B,i}\|_F}.
\]

As described above, we decompose the overall precoding matrix optimization problem (22) into \( D \) precoding sub-matrix optimization problems, which is summarized in the pseudo-code shown in Algorithm 2.

**Algorithm 2 Hybrid Digital and Analog Precoders Design**

**Require:** \( \mathbf{F}, D, \) and \( N \)

1. for \( 1 \leq i \leq D \) do
2. \( \mathbf{B} = (\mathbf{F}_i^H \mathbf{F}_i)^{\frac{1}{2}}, \mathbf{P} = \mathbf{B}^{-1} \mathbf{F}_i \)
3. \( \mathbf{F}_{R,i} = \frac{1}{\sqrt{N}} e^{j \angle (p)} \)
4. \( \mathbf{F}_{B,i} = \mathbf{F}_{R,i}^H \mathbf{F}_i \)
5. \( \mathbf{F}_{B,i}^\text{norm} = \frac{\mathbf{F}_{B,i}}{\|\mathbf{F}_{R,i} \mathbf{F}_{B,i}\|_F} \)
6. end for

**Return:** \( \mathbf{F}_R \) and \( \mathbf{F}_B \)

**C. Energy Efficiency of mmWave Massive MIMO Systems**

It is well known that the energy efficiency of a communications system is determined by the spectral efficiency as well as the total power consumption. For mmWave MIMO systems, the energy efficiency can be written as

\[
\eta = \frac{C}{P_t + N_t P_{RPA} + N_{PS} P_{PS} + N_i P_{PA}},
\]

where \( P_t \) is the transmitter power, \( N_{PS} \) is the required number of phase shifters of the hybrid precoding structure,
matrix. For example, the vector form of the SVD of \( \tilde{A} = H \) is described in (2) with \( L = 8 \) scattering paths [17]. Meanwhile, the AoD and AoA are assumed to follow a uniform distribution within \([-\frac{\pi}{2}, \frac{\pi}{2}]\) and \([-\pi, \pi]\), respectively. Both the transmitter and receiver are equipped with ULAs, and \( d_t = d_r = \frac{\lambda}{4} \).

Finally, the signal-to-noise (SNR) is given as \( \text{SNR} = \frac{\rho}{\sigma^2} \), and the average signal-to-noise ratio (SNR) is averaged over 1000 channel realizations.

### V. Simulation Results

This section presents simulation results on spectral efficiency and energy efficiency to demonstrate the performance of our proposed hybridly connected structure. Meanwhile, we compare the performance of our proposed algorithm using a fixed stream distribution with some recently proposed algorithms for the fully and partially connected structures. In our simulations, we consider the narrow-band channel model described in (2) with \( L = 8 \) scattering paths [17]. Meanwhile, the AoD and AoA are assumed to follow a uniform distribution within \([-\frac{\pi}{2}, \frac{\pi}{2}]\) and \([-\pi, \pi]\), respectively. Both the transmitter and receiver are equipped with ULAs, and \( d_t = d_r = \frac{\lambda}{4} \).

Finally, the signal-to-noise (SNR) is given as \( \text{SNR} = \frac{\rho}{\sigma^2} \), and the average signal-to-noise ratio (SNR) is averaged over 1000 channel realizations.

#### A. Spectral Efficiency

Firstly, according to the relationship between the number of streams \( N_s \) and the equipped number of RF chains \( SD \) in the hybridly connected structure as discussed in Section III, two cases will be discussed, i.e., \( N_s = SD \) and \( N_s < SD \).

1) **Case 1** \((N_s = SD)\): In this case, we assume \( N_s = SD = 8 \). Fig. 3 compares the achievable rates of the three HBF structures, where \( N_t \times N_r = 128 \times 32 \) and \( D \in [2, 4] \). As can be observed from the comparison in Fig. 3, the achievable rate of the hybridly connected structure falls between those of the fully and partially connected structures. Meanwhile, the achievable rate of the proposed structure with \( D = 2 \)
outperforms that with $D = 4$. From this point, we know that the less number of sub-arrays the higher achievable rate is, when the total number of RF chains is equal. Furthermore, the performance of our proposed hybrid constraint precoding algorithm is close to that of its unconstraint counterpart for the hybridly connected structure. Fig. 4 also shows the comparison of the achievable rates of the three structures, where $N_t \times N_r = 128 \times 32$, $N_s = DS = 4$, and $D = 2$. Similar observations can be made from Fig. 4 as those in Fig. 3.

2) Case 2 ($N_s < SD$): In this case, we assume the number of sub-arrays is fixed, and the number of RF chains connected by each sub-array varies. The achievable rates with various numbers of RF chains are shown in Fig. 5, where $N_t \times N_r = 128 \times 32$, $N_s = 8 \leq DS$, $D = 4$, and $S \in \{2, 4, 8\}$. It is observed that the achievable rate of optimal hybrid precoding in this case increases with the number of RF chains connected by each sub-array, and the performance of the proposed constraint precoding algorithm approaches that of the optimal unconstraint precoding. Meanwhile, it is noted that the achievable rate of unconstraint precoding for the hybridly connected structure is equal to that of the fully connected structure, when the number of RF chains connected by each sub-array is equal to the number of streams. Fig. 6 depicts the achievable rates when the number of streams $N_s = 4 \leq DS$, where $N_t \times N_r = 128 \times 32$, $D = 2$, and $S \in \{2, 4\}$. Similar conclusions can be made from Fig. 6 as those in Fig. 5.

As discussed previously, we know that the fully and partially connected structures are the special cases of the hybridly connected structure. Next, we evaluate the achievable rate of the proposed algorithm for these special cases.

Fig. 7 compares the achievable rates of the proposed algorithm, OMP algorithm [17], switch-structure HBF algorithm [26], and directional beamforming method [30], [31] for the mmWave MIMO system, where $N_t \times N_r = 128 \times 32$, $N_s = S = N_R$, and $D = 1$. Since the performances of the OMP and directional beamforming algorithms are affected by the number of paths of the channel, simulation results with various path numbers are presented in Fig. 7. As can be observed from Fig. 7(a), the proposed algorithm outperforms the other algorithms when $N_s = 2$ and $L = 5$. It can be inferred from Fig. 7(b) that the proposed algorithm outperforms the switch-structure HBF algorithm and directional beamforming algorithm, and is slightly inferior to the OMP algorithm when $N_s = 4$ and $L = 5$. Similar conclusions can be drawn from Fig. 7(c) as those from Fig. 7(a) when $N_s = 4$ and $L = 8$. It is worth noting that the performance of the proposed algorithm is extremely close to the optimal unconstrained precoding in all simulation configurations, and is much more robust than the other comparative algorithms. This implies the proposed algorithm not only works effectively in the hybridly connected structure, but is also applicable to the fully connected structure.

Fig. 8 compares the achievable rates of the proposed algorithm with those of the SIC-based HBF algorithm [25] and switch-structure HBF algorithm [26] in the mmWave MIMO system, where $N_t \times N_r = 128 \times 32$, $N_s = DS = N_R = 4$, and $D = 4$. As can be observed from Fig. 8, the proposed algorithm far outperforms the switch-architecture HBF algorithm, and also surpasses the SIC-based HBF algorithm. Moreover, the proposed algorithm is within a small gap from the performance of the optimal unconstrained precoding, which implies the proposed algorithm works in both the hybridly and partially connected structures.
Fig. 7. Comparison of the achievable rates of the comparative algorithms for an $N_t \times N_r = 128 \times 32$ mmWave MIMO system with $D = 1$, and $L = 5$ and $L = 8$.

Fig. 9 compares the achievable rates of different precoding algorithms against the number of RF chains, where $N_t \times N_r = 288 \times 32$, $N_s = 8$, $D = 4$ and SNR = 0 dB. As can be observed from the figure, the achievable rate of the proposed algorithm for the hybridly connected structure increases with the number of RF chains. When $DS = 32$, the achievable rate of the proposed algorithm is very close to that of the fully digital precoding algorithm. Moreover, we observe that the performance of the OMP algorithm can approach that of the fully digital precoding algorithm, when the number of RF chains increases. Meanwhile, it is noted that the performance of the SIC-based precoding algorithm almost does not vary with the number of RF chains.

B. Energy Efficiency

In this part, we will compare the energy efficiency of the three HBF structures according to (33), which is shown in Fig. 10. The simulation parameters are the same as in Fig. 9, and $P_t = 10W$, $P_{PS} = 10mW$, $P_{PA} = 100mW$, and $P_R = 100mW$ [46]. As can be seen from the Fig. 10, the energy efficiency of the hybridly connected structure outperforms those of the fully and partially connected structures in all the case. It is worth pointing out that for the hybridly connected structure, when the number of RF chains is more than the minimum number of RF chains in the fully and partially connected structures, it is still more energy efficient. For example, for the hybridly connected structure with $SD = 32$, the energy efficiency is 0.71 bps/Hz/W, whilst for the fully and partially connected structures with $N_R = 8$, the energy efficiencies are 0.69 bps/Hz/W and 0.65 bps/Hz/W, respectively. Moreover, the energy efficiency of the hybridly connected structure increases...
where $N$ grows more slowly as shown in Fig. 9. When $N$ increases in the number of RF chains, the spectral efficiency decreases with the increase of the number of RF chains. Therefore, the energy efficiency slowly decreases with the increase of the number of RF chains.

first and then decreases slightly with the increase in the number of RF chains. When $SD = 16$, the energy efficiency is greatest, which provides a reference for the design of the actual system. For the fully connected structure, with the increase in the number of RF chains, the spectral efficiency grows more slowly as shown in Fig. 9. When $N_R \geq 26$, the spectral efficiency nearly stops changing with the increase of the number of the RF chains. Therefore, the energy efficiency rapidly reduces with the increase of the number of RF chains as shown in Fig. 10. For the partially connected structure, the spectral efficiency almost does not improve with the increase of the number of RF chains as can be seen from Fig. 9. Meanwhile, the required number of phase shifters is the same, which is equal to the number of antennas $N_t$ at the transmitter as shown in Table 1. Therefore, the energy efficiency slowly decreases with the increase of the number of RF chains.

### C. Sensitivity

To demonstrate the sensitivity of the proposed precoder to the channel estimation error, the achievable rate of the proposed algorithm with various levels of channel estimation errors are presented.

Fig. 11 plots the achievable rates of the proposed algorithm with various numbers of antenna at the transmitter and receiver, where $N_t = SD = 4$, $D = 2$, $S = 2$, and perfect CSI and imperfect CSI with different values of $\zeta$ are considered. As can be observed from Fig. 11, in the case of the same channel accuracy, the larger the number of antennas at the transmitter and receiver, the less the achievable rate loss, and more robust to channel estimation errors. This observation is consistent with the conclusion based upon our theoretical analysis.

### VI. Conclusions

In this paper, we proposed a hybridly connected structure alongside a hybrid digital and analog precoders design in mmWave MIMO systems. In accordance with the structure of the optimal hybrid precoding matrix, the maximum achievable rate optimization problem was decomposed into a series of sub-rate optimization problems for each stream, since each column of the hybrid precoding matrix corresponds to the precoding vector of one stream. Meanwhile, the optimal hybrid precoding sub-matrix can be achieved according to the SVD of the corresponding channel matrix, and the mutual contribution of the sub-matrices can be eliminated using SIC. Furthermore, according to the factorization of the corresponding optimal precoding sub-matrix of each sub-array, the near-optimal hybrid digital and analog precoders were designed to minimize the Euclidean distance with the optimal precoding sub-matrix.

Finally, simulation results were presented to show that the hybridly connected structure falls in between the partially connected structure and the fully connected structure in the sense of spectral efficiency. With an increasing number of RF chains, the spectral efficiency can approach that of the fully connected structure. In terms of energy efficiency, the hybridly connected structure was shown to outperform its partially connected and fully connected counterparts. Meanwhile, the proposed algorithm is insensitive to channel estimation errors.

### REFERENCES


REFERENCES

Didi Zhang received the B.E. degree in electronic and information engineering from the North College of Beijing University of Chemical Technology, Langfang, China, in 2012, and the M.S. degree in information and communication engineering from the North China University of Technology, Beijing, China, in 2015. He is currently pursuing the Ph.D. degree in information and communication engineering from the Beijing University of Posts and Telecommunications, Beijing. His research interests include massive MIMO, mmWave communications, wireless communications, and signal processing.

Yafeng Wang (S’00–M’03–SM’09) received the B.Sc. degree from the Baoji University of Arts and Science in 1997, the M.Eng. degree from the University of Electronic Science and Technology of China in 2000, and the Ph.D. degree from the Beijing University of Posts and Telecommunications in 2003.

In 2008, he was a Visiting Scholar with the Faculty of Engineering and Surveying, University of Southern Queensland, Australia. He is currently a Professor of electronic engineering with the School of Information and Telecommunications, Beijing University of Posts and Telecommunications. He leads the Broadband Mobile Communication Engineering Laboratory, which is one of Zhongguancun Science Park Open Laboratories. He has authored or co-authored over 100 peer-reviewed journal and conference papers. His research mainly focuses on wireless communications and information theory.

Xuehua Li received the Ph.D. degree in telecommunications engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2008. She is currently a Professor and the Deputy Dean of the School of Information and Communication Engineering with Beijing Information Science and Technology University, Beijing. She is a Senior Member of the Beijing Internet of Things Institute. Her research interests are in the broad areas of communications and information theory, particularly the Internet of Things, and coding for multimedia communications systems.

Wei Xiang (S’00–M’04–SM’10) received the B.Eng. and M.Eng. degrees in electronic engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 1997 and 2000, respectively, and the Ph.D. degree in telecommunications engineering from the University of South Australia, Adelaide, Australia, in 2004.

From 2004 to 2015, he was with the School of Mechanical and Electrical Engineering, University of Southern Queensland, Toowoomba, Australia. He is currently the Founding Professor and the Head of the Discipline of Internet of Things Engineering with the College of Science and Engineering, James Cook University, Cairns, Australia. He has authored or co-authored over 200 peer-reviewed journal and conference papers. His research interests are in the broad areas of communications and information theory, particularly the Internet of Things, and coding and signal processing for multimedia communications systems. He is an Elected Fellow of the IET and Engineers Australia. He received the TNQ Innovation Award in 2016, and was a finalist for 2016 Pearcey Queensland Award. He was a co-recipient of three best paper awards at 2015 WCSP, 2011 IEEE WCNC, and 2009 ICWMC. He has been awarded several prestigious fellowship titles. He was named a Queensland International Fellow (2010–2011) by the Queensland Government of Australia, an Endeavour Research Fellow (2012–2013) by the Commonwealth Government of Australia, a Smart Futures Fellow (2012–2015) by the Queensland Government of Australia, and a JSPS Invitational Fellow jointly by the Australian Academy of Science and Japanese Society for Promotion of Science (2014–2015). He is the Vice Chair of the IEEE Northern Australia Section. He was an Editor of the IEEE COMMUNICATIONS LETTERS (2015–2017), and is an Associate Editor of Telecommunications Systems (Springer). He has severed in a large number of international conferences in the capacity of General Co-Chair, TPC Co-Chair, Symposium Chair, and so on.