A New Design of Codebook for Hybrid Precoding in Millimeter-Wave Massive MIMO Systems

Gang Liu [1,2], Honggui Deng [1,*], Kai Yang [1], Zaoxing Zhu [1], Jitai Liu [1] and Hu Dong [2]

1   School of Physics and Electronics, Central South University, Lushan South Road, Changsha 410083, China; 162201003@csu.edu.cn (G.L.); yangkai.rjsb@vivo.com (K.Y.); csuzzx970112@csu.edu.cn (Z.Z.); Liujitai@csu.edu.cn (J.L.)
2   College of Information Science and Engineering, Changsha Normal University, Teli Road, Changsha 410100, China; wjd3203@smail.hunnu.edu.cn
*   Correspondence: denghonggui@csu.edu.cn

Abstract: The precoding scheme based on codebooks is used to save the same set of codebook in advance at the transmitter and the receiver, then, the receiver selects the most appropriate precoding matrix from codebooks according to different channel state information (CSI). Therefore, the design of codebook plays an important role in the performance of the whole scheme. The symmetry-based hybrid precoder and combiner is a highly energy efficient structure in the millimeter-wave massive multiple-input multiple-output (MIMO) system, but at the same time, it also has the problems of high bit error rate and low spectral efficiency. In order to improve the spectral efficiency, we formulate the codebook design as a joint optimization problem and propose an iteration algorithm to obtain the enhanced codebook by combining the compressive sampling matching pursuit (CoSaMP) algorithm with the dictionary learning algorithm. In order to prove the validity of the proposed algorithm, we simulate and analyze the change of the spectral efficiency of the algorithm with the signal-to-noise ratio (SNR) and the number of radio frequency (RF) chains of different precoding schemes. The simulation results demonstrate that the spectral efficiency of the algorithm is obviously outstanding compared with that of the OMP-based joint codebook algorithm and the hybrid precoding algorithm with quantization algorithm under low SNR and different numbers of RF chains. Particularly, when SNR is lower than 0 dB, the proposed algorithm performs very close to the optimal unconstrained precoding algorithm.

Keywords: massive MIMO; millimeter wave; hybrid precoding; codebook

1. Introduction

Millimeter-wave massive MIMO, which is very suitable for 5G wireless communication transmission, has been extensively studied in the past ten years [1]. Although digital precoding can achieve excellent performance in the system, the digital precoding consumes a lot of energy and the hardware cost is quite expensive, owing to the requirement that each antenna element needs to be connected to one dedicated RF chain, which is impractical in the mmWave massive MIMO systems. In order to overcome these obstacles, we utilize hybrid precoding techniques at the cost of reduced spectral efficiency. Therefore, it is vital to design a hybrid precoding algorithm that can promote the spectral efficiency for mmWave MIMO systems.

Over the past decade, hybrid precoding technology in the mmWave massive MIMO system has been diffusely researched. By regarding the mmWave channel as sparse channel, the construction of hybrid precoding is transformed into the sparse matrix reconstruction in [2], although the hybrid precoding matrix is easy to construct, the adopted orthogonal matching pursuit (OMP) algorithm causes error accumulation, which causes the decrease of system spectral efficiency. The transceiver is designed for energy efficiency maximization in [3] by dividing the joint design of the hybrid precoder and the combiner into two

subproblems, and further, using the OMP algorithm to design the hybrid precoder, then improving the design of the hybrid precoder based on random error modeling in order to adapt to the situation of imperfect CSI. With the aid of the internal connection between the analog codebook and digital codebook, an iterative algorithm is designed jointly by digital codebooks and analog codebooks in [4] based on the dictionary learning algorithm and reconstruction algorithm, which can be applied to the phase shifter with arbitrary resolution, but the algorithm has not improved the spectral efficiency. According to the different connection modes between radio frequency (RF) chains and antennas, the hybrid precoding scheme is mainly divided into full connection, partial connection, and hybrid connection. In [5,6], a hybrid precoding scheme based on the partial connection structure is proposed, which reduces the number of phase shifters by connecting each RF chain with partial antennas, thus, effectively reducing the hardware complexity. However, the hybrid precoding based on partial connection structure will bring some loss of spectral efficiency. Hybrid-connected hybrid precoding is a trade-off between full connection and partial connection. The antennas are divided into multiple subarrays by RF chain, and each subarray is fully connected. Reference [7] proposes a successive interference cancellation hybrid precoding scheme for hybrid-connected hybrid precoding. According to the relationship between the transmitted data stream and the number of RF chains, the form of digital precoding matrix is designed first, and then the analog precoding matrix is decomposed into a submatrix by using the block diagonal characteristics of the analog precoding matrix. Finally, the hybrid precoding matrix is obtained, which is better than the partially connected hybrid precoding scheme. Compared with partial connection and hybrid connection, fully-connected hybrid precoding has better spectral efficiency performance because all RF chains are connected to each antenna at the transmitter, which can provide higher degree of freedom for the transmitted signal. It is closer to the full digital precoding scheme, but its hardware complexity is high. Therefore, in recent years, some researchers have explored the use of machine learning and deep learning methods to realize the hybrid precoding algorithm based on full connection structure, in order to reduce the hardware cost. For example, in [8], they propose a novel method based on deep learning, which designs the hybrid beamforming vector by jointly optimizing the channel measurement vector, and achieves better spectral efficiency. However, the training cost is high for each update, that is to say, the computational complexity is high. The digital precoding codebook and analog precoding codebook are designed separately to reduce the complexity of the algorithm in [9,10], but the relationship between them is not considered and exploited, resulting in the loss of system spectral efficiency. The hybrid precoding algorithm proposed in [11] reduces the complexity of the algorithm by avoiding singular value decomposition (SVD) and matrix inversion but at the cost of spectral efficiency. The hybrid precoding algorithm in [12,13] is only applicable to the frequency division duplexing (FDD) system with known channel state information (CSI), but the analog codebook construction scheme is no longer applicable in the scenario where the codebook size and phase shifter resolution are determined so that application scenarios are not flexible enough.

Inspired by the aforementioned algorithms, we propose an improved iterative algorithm of the hybrid precoding codebook to improve the spectral efficiency of a system. Firstly, we convert the design problem of hybrid precoding codebook into a minimization problem, which aims to minimize the Euclidean distance between the designed hybrid precoder and the optimal unconstrained precoder. Further, we apply the CoSaMP algorithm to reconstruct the digital precoding matrix. Then, with the aid of channel sparsity in mmWave Massive MIMO System, after taking the relationship between analog and digital precoders into consideration, we jointly generate the hybrid codebook by combining reconstruction algorithms with dictionary learning algorithms. Finally, the qualified size digital precoding codebook is generated using the clustering algorithm. Simulation results show that the hybrid precoding matrix generated by the proposed algorithm is obviously better than

the OMP-based codebook [4] algorithm and the hybrid precoding with quantization [1] algorithm in low-SNR regimes and with different numbers of RF chains.

The remainder of this paper is organized as follows. The extended Saleh–Valenzuela [14] channel model and the system model of hybrid precoding based on the mmWave massive MIMO system is introduced and the objective function is also derived in Section 2. The iterative hybrid precoding codebook algorithm is proposed to improve the spectral efficiency in Section 3. A simulation analysis of the algorithm and comparative discussion with the existing algorithm are presented in Section 4. Finally, we conclude this paper in Section 5.

Notations: For a matrix $\mathbf{A}$, the transpose and conjugate transpose of $\mathbf{A}$ are represented by $\mathbf{A}^T$ and $\mathbf{A}^H$; $|\mathbf{A}|$ denotes the determinant; $\|\mathbf{A}\|_F$ and $\|\mathbf{A}\|_0$ denote the Frobenius norm and zero norm of $\mathbf{A}$; $diag(\mathbf{A})$ is a vector formed by the diagonal elements of $\mathbf{A}$; $\mathbf{A}(:,n)$ and $\mathbf{A}(m,:)$ denote the nth column and mth row of $\mathbf{A}$, respectively; $\mathbf{A}(m,n)$ denotes the mth row and nth column element of $\mathbf{A}$.

2. System Model

In this section, the mmWave massive MIMO system model, channel model, and joint codebooks design scheme are established.

2.1. Channel Model

The overall system architecture is shown in Figure 1. As shown in the single-user mmWave massive MIMO system, we consider a fully-connected structure in the system in which each radio frequency (RF) chain will be connected to all antennas through phase shifters. In addition, the antenna array will adopt a uniform linear array (ULA) configuration. In this system, the transmitter and receiver are equipped with N_t and N_r antennas, and N_t^{RF} and N_r^{RF} RF chains, where N_t and N_r denote the number of antennas for the transmitter and receiver, respectively, and N_t^{RF} and N_t^{RF} denote the number of RF chains for the transmitter and receiver, respectively. For the hybrid precoding system, the number of data streams is N_s, where $N_s \leq N_t^{\mathrm{RF}} \leq N_t$, $N_s \leq N_r^{\mathrm{RF}} \leq N_r$. Unlike traditional digital precoding, the hybrid precoder is composed of a digital precoder $\mathbf{F}_{\mathrm{BB}}$ and an analog precoder $\mathbf{F}_{\mathrm{RF}}$, where $\mathbf{F}_{\mathrm{BB}}$ denotes the digital precoding matrix of size $N_t^{\mathrm{RF}} \times N_s$ and $\mathbf{F}_{\mathrm{RF}}$ denotes the analog precoding matrix of size $N_t \times N_t^{\mathrm{RF}}$. So, after passing the hybrid precoder, the transmitted signal is $\mathbf{x} = \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\mathbf{s}$, where $\mathbf{s}$ denotes the data stream's vector satisfying $\mathbb{E}[\mathbf{ss}^H] = \frac{1}{N_s}\mathbf{I}_{N_s}$, where $\mathbf{I}_{N_s}$ denotes the identity matrix of size $N_s \times N_s$. With the limitation of the total power in the transmitter, the hybrid precoder power needs to satisfy $\|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_F^2 = N_s$. At the receiver, there are also a $N_r^{\mathrm{RF}} \times N_s$ digital combining matrix $\mathbf{W}_{\mathrm{BB}}$ and a $N_r \times N_r^{\mathrm{RF}}$ analog combining matrix $\mathbf{W}_{\mathrm{RF}}$, so the received signal vector $\mathbf{y} \in \mathbb{C}^{N_r \times 1}$ can be expressed as [14]

$$\mathbf{y} = \sqrt{\rho}\mathbf{W}_{\mathrm{BB}}^H\mathbf{W}_{\mathrm{RF}}^H\mathbf{H}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\mathbf{s} + \mathbf{W}_{\mathrm{BB}}^H\mathbf{W}_{\mathrm{RF}}^H\mathbf{n}, \tag{1}$$

where ρ is the average received power, $\mathbf{H}$ of size $N_r \times N_t$ is the channel matrix, and $\mathbf{n}$ denotes the noise that follows the distribution $\mathcal{CN}(0, \sigma_n^2)$.

The features of mmWave transmission are the high free-space path loss and the large-scale antenna arrays, resulting in limited spatial selectivity or scattering [14] and high correlation of antennas [15]. For mmWave channel modeling, many of the classical statistical fading distributions used in traditional MIMO scenarios are obviously not applicable to the characteristic of mmWave transmission environments. In order to better utilize the feature in mmWave channels, we present a narrowband clustered channel model, according to the extended Saleh–Valenzuela model [14,15]. Therefore, the discrete-time narrowband clustered channel model can be expressed as [14–16]

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{N_{\mathrm{cl}} N_{\mathrm{ray}}}} \sum_{i=1}^{N_{\mathrm{cl}}} \sum_{l=1}^{N_{\mathrm{ray}}} \alpha_{il} \mathbf{a}_{\mathrm{r}}(\phi_{il}^{\mathrm{r}}) \mathbf{a}_{\mathrm{t}}(\phi_{il}^{\mathrm{t}})^H, \tag{2}$$

where $\mathbf{H}$ is the channel matrix, α_{il} is the complex gain of the l-th path in the i-th cluster that is subjected to the complex Gaussian distribution $\mathcal{CN}(0, \sigma_{\alpha,i}^2)$, where $\sum_{i=1}^{N_{cl}} \sigma_{\alpha,i}^2 = \sqrt{\frac{N_t N_r}{N_{cl} N_{ray}}}$ and N_{cl} and N_{ray} denote the number of clusters and the number of paths in each cluster, respectively. ϕ_{il}^r and ϕ_{il}^t are the angles of arrival and departure (AoAs and AoDs, respectively) of the lth path in the ith cluster, respectively. $\mathbf{a}_r(\phi_{il}^r)$ and $\mathbf{a}_t(\phi_{il}^t)$, respectively, denote the antenna response array at the receiver and transmitter, which are given by [17]

$$
\begin{aligned}
\mathbf{a}_t(\phi_{il}^t) &= \frac{1}{\sqrt{N_t}} \left[1, e^{jkdsin(\phi_{il}^t)}, \cdots, e^{j(N_t-1)kdsin(\phi_{il}^t)} \right]^T, \\
\mathbf{a}_r(\phi_{il}^r) &= \frac{1}{\sqrt{N_r}} \left[1, e^{jkdsin(\phi_{il}^r)}, \cdots, e^{j(N_r-1)kdsin(\phi_{il}^r)} \right]^T,
\end{aligned}
\tag{3}
$$

where $k = 2\pi/\lambda$, λ is the wavelength and $d = \frac{\lambda}{2}$ is the interelement spacing. In addition, ϕ_i is uniformly distributed over $[0, 2\pi)$ with the Laplacian distribution.
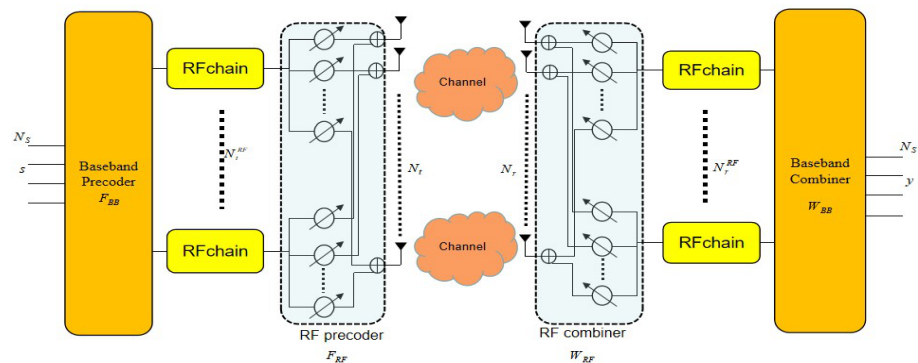
Figure 1. Simplified block diagram of the mmWave hybrid precoding system with the fully-connected transceiver architecture.

2.2. Problem Formulation

Based on Shannon's theorem, the spectral efficiency of the system in Figure 1 can be formulated as follows:

$$
R = \log_2\left(\left| \mathbf{I}_{N_s} + \frac{\rho}{N_s} \mathbf{R}_n^{-1} \mathbf{W}_{BB}^H \mathbf{W}_{RF}^H \mathbf{H} \mathbf{F}_{RF} \mathbf{F}_{BB} \mathbf{F}_{BB}^H \mathbf{F}_{RF}^H \mathbf{H}^H \mathbf{W}_{RF} \mathbf{W}_{BB} \right|\right),
\tag{4}
$$

where $\mathbf{R}_n = \sigma_n^2 \mathbf{W}_{BB}^H \mathbf{W}_{RF}^H \mathbf{W}_{RF} \mathbf{W}_{BB}$ denotes the combined noise covariance matrix.

We need to jointly optimize the four variables $(\mathbf{F}_{RF}, \mathbf{F}_{BB}, \mathbf{W}_{RF}, \mathbf{W}_{BB})$ in (4) to maximize spectral efficiency, but they are hard to optimize together in practice [18,19]. By decoupling the joint optimization problem at the transmitter and receiver, the problem of maximizing spectral efficiency can be transformed into a problem of maximizing mutual information in [1], which can be expressed as follows:

$$
I(\mathbf{F}_{RF}, \mathbf{F}_{BB}) = \log_2\left(\left| \mathbf{I}_{N_s} + \frac{\rho}{N_s \sigma_n^2} \mathbf{H} \mathbf{F}_{RF} \mathbf{F}_{BB} \mathbf{F}_{BB}^H \mathbf{F}_{RF}^H \mathbf{H}^H \right|\right),
\tag{5}
$$

where $I(\mathbf{F}_{RF}, \mathbf{F}_{BB})$ denotes the amount of mutual information that can be maximized by optimizing the design of the hybrid precoder $(\mathbf{F}_{RF}, \mathbf{F}_{BB})$; the optimization of the analog and digital precoding matrix is expressed as follows:

$$
\begin{aligned}
(\mathbf{F}_{RF}^{opt}, \mathbf{F}_{BB}^{opt}) &= \arg\max_{\mathbf{F}_{RF}, \mathbf{F}_{BB}} I(\mathbf{F}_{RF}, \mathbf{F}_{BB}), \\
\text{s.t.} \quad &\mathbf{F}_{RF} \in \mathcal{F}_{RF}, \\
&\|\mathbf{F}_{RF} \mathbf{F}_{BB}\|_F^2 = N_s,
\end{aligned}
\tag{6}
$$

where $\mathcal{F}_{\mathrm{RF}}$ is the feasible set whose elements are satisfied with the constraint of constant modulus and it is hard to find a general solution to work out (6) under the nonconvex constraint $\mathbf{F}_{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}$. Refs. [1,4] proved that maximizing the above amount of mutual information can be converted to the minimization problem, which aims to minimize the Euclidean distance between the designed hybrid precoder and the optimal unconstrained precoder. So, the design problem of hybrid precoders can be further reformulated as the follows:

$$(\mathbf{F}_{\mathrm{RF}}^{\mathrm{opt}}, \mathbf{F}_{\mathrm{BB}}^{\mathrm{opt}}) = \arg \min_{\mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{BB}}} \left\| \mathbf{F}_{\mathrm{opt}} - \mathbf{F}_{\mathrm{RF}} \mathbf{F}_{\mathrm{BB}} \right\|_F,$$

$$\text{s.t.} \quad \mathbf{F}_{\mathrm{RF}}(m, n) \in \mathcal{F}_{\mathrm{RF}}, \forall m, n, \tag{7}$$

$$\left\| \mathbf{F}_{\mathrm{RF}} \mathbf{F}_{\mathrm{BB}} \right\|_F^2 = N_{\mathrm{s}},$$

where $\mathbf{F}_{\mathrm{opt}}$ is the optimal unconstrained precoder that can theoretically achieve the best spectral efficiency. Additionally, $\mathbf{F}_{\mathrm{opt}} = \mathbf{V}(:, 1 : N_{\mathrm{s}})$, where $\mathbf{V}$ is the right singular matrix of the channel matrix $\mathbf{H} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H$ and can be obtained by performing singular value decomposition (SVD). In addition, when using a $q$bit resolution phase shifter, the phase shifter can achieve 2^q phase angles, so $\mathcal{F}_{\mathrm{RF}}$ can be obtained by $\mathcal{F}_{\mathrm{RF}} = \left\{ 1, \omega, \omega^2, \cdots, \omega^{2^q - 1} \right\}$, $\omega = e^{j2\pi/2^q}$.

3. Proposed Joint Codebooks Design Algorithm

In order to save feedback overhead, the hybrid precoder, calculated at the receiver, needs to feed back to the transmitter after quantization in the FDD system. Consequently, it is vital to design the appropriate hybrid precoder codebook to improve the system performance. According to CoSaMP algorithm, a new joint codebooks design algorithm is introduced and combined with the dictionary learning algorithm to generate the digital precoding matrix and analog precoding matrix at the same time; this algorithm has improved system spectral efficiency compared with existing algorithms.

In this paper, we assume that the size of the analog and digital codebook are N_1 and N_2, respectively. Under the condition that the size of the analog codebook is determined, the number of phase angles that the corresponding phase shifter can achieve is also determined as N_1. The nonzero elements of the analog codebook are contained from (3) and the angles are uniformly distributed in the range of $[0, 2\pi]$, therefore, the analog codebook can be written as $\mathcal{F}_{\mathrm{A}} = \left[\mathbf{F}_1^{\mathrm{A}}, \mathbf{F}_2^{\mathrm{A}}, \cdots, \mathbf{F}_{N_1}^{\mathrm{A}} \right] = \left[\mathbf{a}_t \left(\frac{2\pi}{N_1} \right), \mathbf{a}_t \left(\frac{2\pi}{N_1} 2 \right), \cdots, \mathbf{a}_t (2\pi) \right]$. The expression of (7) is equivalent to the problem of sparse signal recovery, except that the variables are different [1]. When the initial analog codebook is determined, the joint optimization problem of $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{F}_{\mathrm{BB}}$ can be converted into the reconstruction problem of the sparse constraint matrix with only one variable, so the optimal problem (7) can be rewritten as the following equivalent problem:

$$\widetilde{\mathbf{F}}_{\mathrm{BB}}^{\mathrm{opt}} = \arg \min_{\widetilde{\mathbf{F}}_{\mathrm{BB}}} \left\| \mathbf{F}_{\mathrm{opt}} - \mathbf{F}_{\mathrm{A}} \widetilde{\mathbf{F}}_{\mathrm{BB}} \right\|_F,$$

$$\text{s.t.} \left\| \mathrm{diag} \left(\widetilde{\mathbf{F}}_{\mathrm{BB}} \widetilde{\mathbf{F}}_{\mathrm{BB}}^H \right) \right\|_0 = N_{\mathrm{t}}^{\mathrm{RF}}, \tag{8}$$

$$\left\| \mathbf{F}_{\mathrm{A}} \widetilde{\mathbf{F}}_{\mathrm{BB}} \right\|_F^2 = N_{\mathrm{s}},$$

where $\left\| \mathrm{diag} \left(\widetilde{\mathbf{F}}_{\mathrm{BB}} \widetilde{\mathbf{F}}_{\mathrm{BB}}^H \right) \right\|_0 = N_{\mathrm{t}}^{\mathrm{RF}}$ is a sparsity constraint and the sparsity of the digital precoding matrix is fixed, which equals to the number of RF chains at the transmitter. When there are only $N_{\mathrm{t}}^{\mathrm{RF}}$ nonzero rows in the $\widetilde{\mathbf{F}}_{\mathrm{BB}}$, there are correspondingly only $N_{\mathrm{t}}^{\mathrm{RF}}$ columns selected in the analog codebook $\mathbf{F}_{\mathrm{A}}$. Therefore, the digital precoder $\mathbf{F}_{\mathrm{BB}}$ is the composition of the nonzero $N_{\mathrm{t}}^{\mathrm{RF}}$ row of $\widetilde{\mathbf{F}}_{\mathrm{BB}}$, and the analog precoder $\mathbf{F}_{\mathrm{RF}}$ is composed of the corresponding $N_{\mathrm{t}}^{\mathrm{RF}}$ column in the $\mathbf{F}_{\mathrm{A}}$.

We can obtain the channel sample matrix from the channel model and then obtain the corresponding optimal unconstrained matrix, so according to Formula (8), we can get the

corresponding analog precoder and digital precoder. The paper utilizes dictionary learning to construct analog codebooks and digital codebooks, i.e., the dictionary is firstly fixed to update the sparse matrix, then the sparse matrix is fixed to update the dictionary, the iteration steps are repeated until the dictionary and sparse matrix are optimal. According to (8), the $\widetilde{\mathbf{F}}_{BB}$ can be calculated by the initial codebook $\mathbf{F}_A$; by fixing the value of $\widetilde{\mathbf{F}}_{BB}$, the design of analog codebook can be expressed as

$$
\begin{aligned}
\mathbf{F}_A^{opt} =& \arg\min_{\mathbf{F}_A} \sum_{i=1}^{N_{tr}} \left\| \mathbf{F}_{opt}(i) - \mathbf{F}_A\widetilde{\mathbf{F}}_{BB}(i) \right\|_F, \\
\text{s.t.} \quad & \mathbf{F}_A(m,n) \in \mathcal{F}_{RF}, \forall m,n, \\
& \left\| \mathbf{F}_A\widetilde{\mathbf{F}}_{BB}(i) \right\|_F^2 = N_s, \forall i,
\end{aligned}
\tag{9}
$$

where N_{tr} denotes the number of channel samples, and $\mathbf{F}_{opt}(i)$ and $\widetilde{\mathbf{F}}_{BB}(i)$ denote the optimal unconstrained precoder and digital precoder, respectively, for the i-th channel sample. It is hard to find a general solution to solve (9) with the nonconvex constraint $\mathbf{F}_A(m,n) \in \mathcal{F}_{RF}$. Therefore, we propose an iterative algorithm to achieve the local optimal solution of the problem.

The proposed algorithm is specifically presented in Algorithm 1, which can be mainly divided into three step. The first stage is step 1, several channel matrix samples are generated offline through the channel model in Formula (2), then, the optimal unconstrained precoding matrix can be calculated by SVD according to each channel matrix and formed into a training set; the second stage includes step 2 to step 14, the sparse matrix (the digital precoding matrix $\widetilde{\mathbf{F}}_{BB}(i)$) reconstructed by the improved OMP (orthogonal matching pursuit) algorithm is not accurate enough in [4] because the selected columns will always be retained in each iteration, so the columns with larger errors cannot be replaced in subsequent processing, resulting in insufficient accuracy of the reconstructed matrix. To solve this problem, we propose a hybrid precoding codebook generation algorithm, according to a combination of dictionary learning and the reconstruction algorithm, which is separated into two parts. In the first part, we utilize the proposed algorithm to find the analog codebook $\mathbf{F}_A$ and the digital codebook $\mathbf{D}$ of size N_{tr}. In each iteration, we select multiple column vectors from the initial analog codebook instead of a single-column vector to form an analog precoding matrix, and then use the least squares method to figure out the digital precoding matrix. Finally, we select columns whose number equals to the number of data stream, from the digital precoding matrix calculated in the previous step, to form the final digital precoding matrix. In this way, the column selected last time can be replaced with a column with a smaller error in the next iteration, avoiding the accumulation of errors, and the specific steps are shown in Steps 3–11. The second part is that the analog and digital codebook will be simultaneously updated and optimized to achieve the maximum system spectral efficiency based on dictionary learning, which aims to further reduce the Euclidean distance between the updated hybrid precoder and the optimal unconstrained precoder. In addition, in the process of updating the analog codebook $\mathbf{F}_A$, we should consider the limitation of the resolution of the phase shifter, so each element of the matrix in the $\mathbf{F}_A$ must be included in $\mathcal{F}_{RF} = \left\{ 1, \omega, \omega^2, \cdots, \omega^{2^q-1} \right\}$ [4].

The third stage includes Step 15; with the requirements of the size of the digital codebook, we can use N_{tr} digital precoding matrices in the digital codebook $\mathbf{D}$ as training samples, and utilize the clustering algorithm to generate a qualified size digital precoding codebook $\mathcal{F}_D = \left[\mathbf{F}_1^D, \mathbf{F}_2^D, \cdots, \mathbf{F}_{N_2}^D \right]$ [20].

Finally, we summarize the whole algorithm flow of the proposed Algorithm 1 in Figure 2.
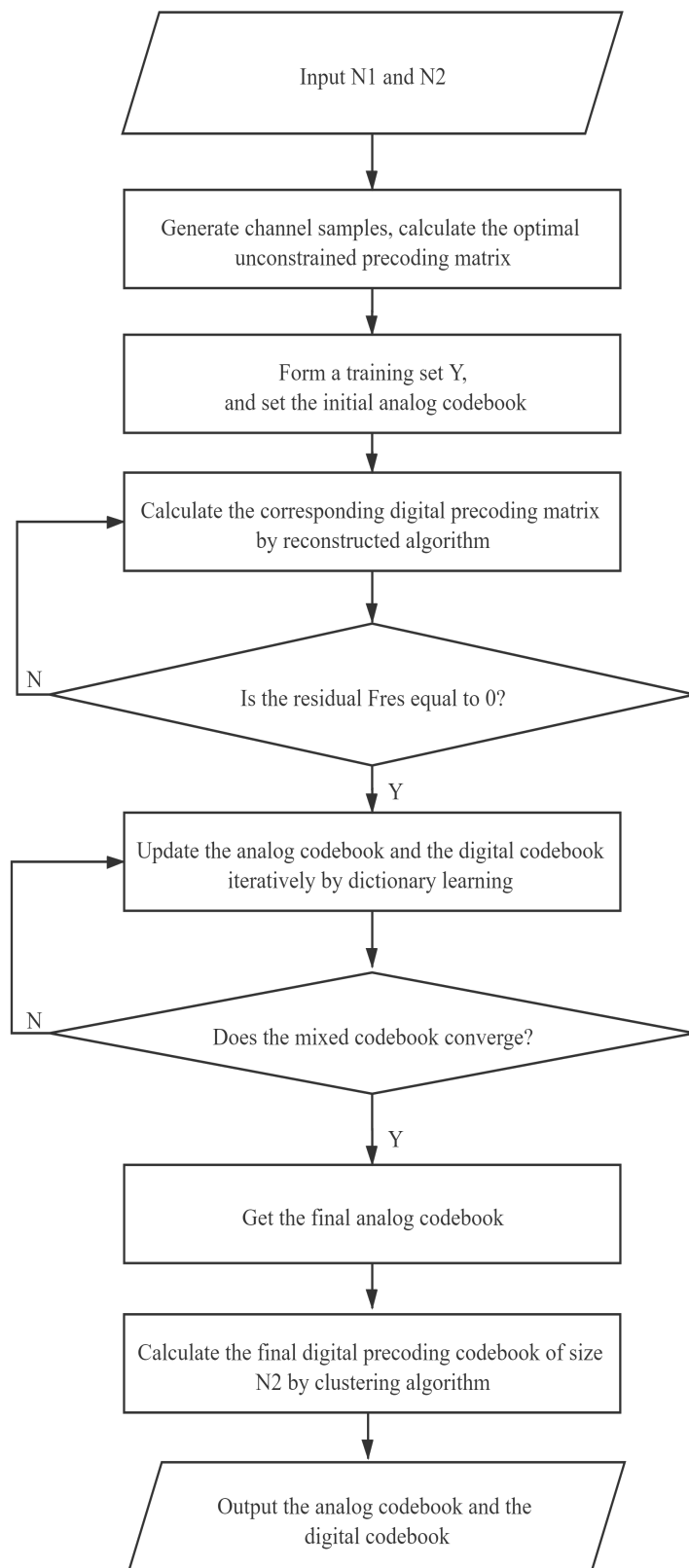
Figure 2. Flow of the proposed joint codebook design algorithm schemes.

Algorithm 1 Hybrid precoding codebook generation algorithm based on dictionary learning and the reconstruction algorithm

---

Input: N_1, N_2

Output: $\mathcal{F}_A, \mathcal{F}_D$

Step1. The channel samples are generated by the channel model in Formula (2), we can find the optimal unconstrained precoding matrix corresponding to each channel matrix and form them into a training set $\mathbf{Y} = [\mathbf{F}_{opt}(1), \mathbf{F}_{opt}(1), \cdots, \mathbf{F}_{opt}(N_{tr})]$, and set the initial simulation codebook $\mathcal{F}_A = \left[\mathbf{a}_t\left(\frac{2\pi}{N_1}\right), \mathbf{a}_t\left(\frac{2\pi}{N_1}2\right), \cdots, \mathbf{a}_t(2\pi)\right]$;

Step2. According to steps 3–12, we can calculate the extended digital precoding matrix $\widetilde{\mathbf{F}}_{BB}(i)$ corresponding to each $\mathbf{F}_{opt}(i)$, the constraints and objective function can be expressed as

$$\widetilde{\mathbf{F}}_{BB}^{opt}(i) = \arg \min_{\widetilde{\mathbf{F}}_{BB}(i)} \left\| \mathbf{F}_{opt}(i) - \mathbf{F}_A \widetilde{\mathbf{F}}_{BB}(i) \right\|_{F},$$

$$\text{s.t.} \quad \left\| \text{diag}\left(\widetilde{\mathbf{F}}_{BB}(i) \widetilde{\mathbf{F}}_{BB}(i)^H \right) \right\|_0 = N_t^{RF}, \forall i, \tag{10}$$

$$\left\| \mathbf{F}_A \widetilde{\mathbf{F}}_{BB}(i) \right\|_F^2 = N_s, \forall i.$$

Step3. Let residual $\mathbf{F}_{res} = \mathbf{F}_{opt}$, Ω is the empty set $\varnothing$;

Step4. Calculate $\boldsymbol{\Phi} = \mathbf{F}_A^H \mathbf{F}_{res}$;

Step5. Select the number of $2N_s$ with the largest value on the diagonal of the matrix $\boldsymbol{\Phi}\boldsymbol{\Phi}^H$ and set the corresponding sequence number θ_1, $\theta_1 = \left\{ k | (\boldsymbol{\Phi}\boldsymbol{\Phi}^H)_{(k,k)} \right\}$;

Step6. Select the corresponding column from the analog codebook FA to compose a matrix $\mathbf{A}_t$, $\mathbf{A}_t = \mathbf{F}_A(:, \theta)$, where $\theta = \Omega \cup \theta_1$;

Step7. The digital precoding matrix $\mathbf{F}_{BB}$ can be calculated by the least squares method, $\mathbf{F}_{BB} = \left(\mathbf{A}_t^H \mathbf{A}_t \right)^{-1} \mathbf{A}_t^H \mathbf{F}_{opt}$;

Step8. Select the number of Ns with the largest value on the diagonal of the matrix $\mathbf{F}_{BB}\mathbf{F}_{BB}^H$ and set the corresponding sequence number θ_2, $\theta_2 = \left\{ q | (\mathbf{F}_{BB}\mathbf{F}_{BB}^H)_{(q,q)} \right\}$;

Step9. Let $\Omega = \theta(:, \theta_2)$, and select the corresponding N_s serial numbers from the set θ;

Step10. Update the residual, $\mathbf{F}_{res} = \mathbf{F}_{opt} - \mathbf{F}_A(:, \theta_2)\mathbf{F}_{BB}(\theta_2, :)$;

Step11. Determine whether the residual is equal to 0; if it equals to 0, enter Step 12; otherwise, return to Step 4 to continue iterating;

Step12. Extend $\mathbf{F}_{BB}(\theta_2, :)$ to $\widetilde{\mathbf{F}}_{BB}(i)$, i.e., the elements are all zero in the rows of non-Ω in $\widetilde{\mathbf{F}}_{BB}(i)$, and the row corresponding to the sequence number in the set Ω is composed of the corresponding elements in $\mathbf{F}_{BB}(\theta_2, :)$;

Step13. Let $\mathbf{X} = \left[\widetilde{\mathbf{F}}_{BB}(1), \widetilde{\mathbf{F}}_{BB}(2), \cdots, \widetilde{\mathbf{F}}_{BB}(N_{tr}) \right]$, we can update $\mathbf{Y}$ by $\mathbf{Y} = \mathbf{F}_A \mathbf{X}$;

Step14. Regarding $\mathbf{Y}$ as the original sample, $\mathbf{F}_A$ as a dictionary, and $\mathbf{X}$ as a sparse matrix, by using the dictionary learning algorithm, the analog codebook $\mathbf{F}_A$ and digital codebook $\mathbf{X}$ can simultaneously update. If the mixed codebook does not converge, the updated simulation codebook $\mathbf{F}_A$ is used as the new initial simulation codebook and we return to Step 2 to continue iteration; if it is satisfied, enter Step 15;

Step15. We can obtain the unextended digital codebook $\mathbf{D}$, $\mathbf{D} = [\mathbf{F}_{BB}(1), \mathbf{F}_{BB}(2), \cdots, \mathbf{F}_{BB}(N_{tr})]$, from the above steps, the final digital precoding codebook $\mathcal{F}_D$ with a codebook size of N_2 can be calculated by using the Lloyd clustering algorithm; the analog precoding codebook is the updated $\mathcal{F}_A$.

---

4. Simulation Results

4.1. Simulation Setup and Explanations

In this section, we simulate the proposed algorithm, varying with different SNR and RF chains, and compare the simulation results with some existing algorithms in terms of system spectral efficiency. The channel model we adopted is the discrete-time narrowband channel model and there are $N_{cl} = 4$ clusters where $N_{ray} = 20$ paths in each cluster. For

the convenience of simulation, we assume that all clusters have the same power, the same azimuth angular spreads of AoDs and AoAs, and the angle expansion is 6^0. The number of channel samples N_{tr} is 5000 and the size of analog precoding codebook size N_1 is equal to the size of the digital precoding codebook $N_2 = 64$. The detailed simulation parameters are shown in Table 1.

Table 1. Simulation parameters.

Parameter	Value
Number of clusters N_{cl}	4
Number of propagation paths per cluster N_{ray}	20
Antenna array deployed	USPA
AoA and AoD cluster angles distribution	Uniform$[0, 2\pi]$
Angular spread	6^0
Number of transmitter antennas N_t	64
Number of receiver antennas N_r	32
Number of of RF chains N_t^{RF}	8
Number of data streams N_s	3

For the convenience of numerical analysis, we give a brief description of the algorithm proposed in this paper and some other algorithms compared with it.

(*a*) **Hybridprecoding algorithm with quantization [1]:** This algorithm is proposed in reference [1]. The algorithm is based on the OMP algorithm for analog codebook design, their digital codebook design directly uses the Grassmannian codebook, and does not use some clustering algorithms to reduce codebook space.

(*b*) **OMP-based Codebook [4]:** This algorithm is proposed in reference [4]. Firstly, the algorithm was based on the improved OMP algorithm for analog codebook design; then, they used the dictionary learning algorithm for digital codebook design; finally, Lloyd clustering algorithm was used for codebook space reduction.

(*c*) **Proposed Codebook:** The algorithm is proposed in this paper. We have elaborated on it in Section 3.

(*d*) **Optimal unconstrained Precoding:** This algorithm is directly based on SVD decomposition of channel **H** to get the optimal unconstrained precoder $\mathbf{F}_{opt}$. Although the algorithm can get the best spectral efficiency theoretically, it is impossible to achieve the optimal precoder without considering the constraints in practice.

4.2. Comparison and Analysis of Spectral Efficiency Performance

Figure 3 indicates that the spectral efficiency achieved by different algorithms varying with SNR and the number of RF chains at the transmitter N_t^{RF} is 8. Figure 3 clearly shows that the spectral efficiency of the proposed codebook design algorithm is higher than that of the OMP-based joint codebook algorithm in [4] and the hybrid precoding algorithm with quantization in [1] in the SNR interval [−5, 30]. The specific simulation results are shown in Table 2. When SNR is 0 dB, the spectral efficiency of the hybrid precoding with quantification algorithm [1] is 14.4082 bps/Hz, the OMP-based codebook algorithm [4] is 15.3462 bps/Hz, the proposed codebook algorithm is 15.8109 bps/Hz, and the optimal unconstrained precoding algorithm is 15.3462 bps/Hz. Compared with the previous two algorithms, the proposed algorithm improves the spectral efficiency by 9.74% and 3.02%, respectively. However, compared with the last algorithm, the proposed algorithm reduces the spectral efficiency by 4.43%. When SNR is 5 dB, the proposed algorithm improves the spectral efficiency by 6.96% and 2.56%, respectively, and reduce by 6.30%. When SNR is 10 dB, the proposed algorithm improves the spectral efficiency by 5.68% and 2.27%, respectively, and reduce by 8.97%. When SNR is between −5 db and 30 dB, although our algorithm is lower than the last algorithm (optimal unconstrained precoding algorithm) in spectral efficiency performance, the last algorithm is executed in an ideal state, which is not realizable. We only compare it as a benchmark. In addition, through the analysis of

the data in Table 2, we can see that compared with the current two better algorithms, our hybrid precoding design algorithm has a significant improvement in spectral efficiency performance when the signal-to-noise ratio (SNR) is low. At the same time, Figure 3 shows that when the SNR is relatively high, the spectral efficiency of our proposed algorithm is not greatly improved compared with the other two algorithms; especially, when the SNR is higher than 25 dB, it almost coincides with the corresponding curves of the two compared algorithms. The reasons are as follows. The algorithms involved in this paper are hybrid precoding codebook design schemes based on the compressed sensing (CS) reconstruction algorithm. The algorithms generally need some prior knowledge of the reconstructed signal, which is some channel state information (CSI) obtained from channel estimation. When SNR is low, the channel estimation result is not accurate enough. The proposed algorithm selects multiple column vectors instead of a single-column vector to form the initial analog codebook at each iteration. When reconstructing the signal, it avoids the error accumulation effect. Compared with other algorithms, the reconstructed signal can be obtained more accurately. So, the spectral efficiency of proposed algorithm is better than other algorithms. However, when SNR is high, the channel estimation result is accurate enough and all algorithms will not accumulate too much error in iterations. This is the reason why the spectral efficiency of the new proposed algorithm is the same as the other previous algorithms for higher SNR regimes.

Table 2. Spectral efficiency of different algorithms with SNR.

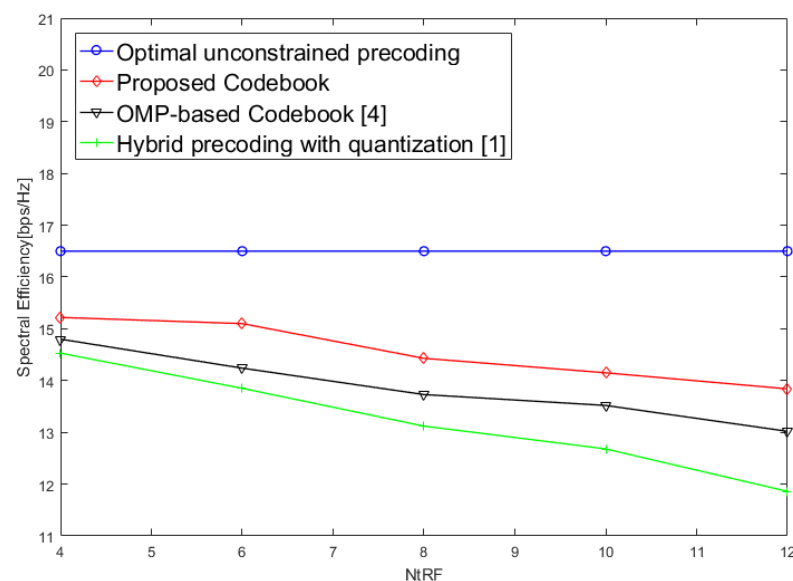| Algorithm	Spectral Efficiency (bps/Hz)		
	SNR = 0 dB	SNR = 5 dB	SNR = 10 dB
Hybrid Precoding with quantization [1]	14.4082	17.5505	20.8505
OMP-based Codebook [4]	15.3462	18.3034	21.5461
Proposed Codebook	15.8109	18.7715	22.0345
Optimal unconstrained precoding	16.5443	20.0332	24.0119
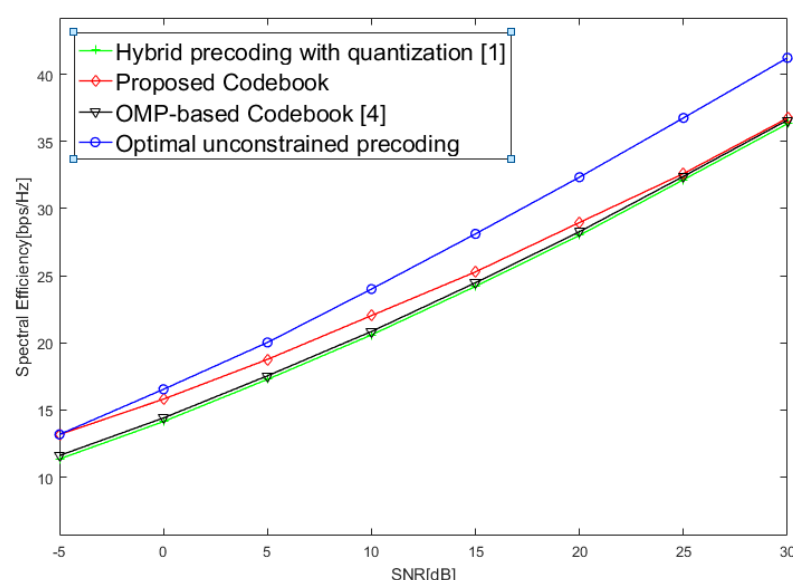
Figure 3. Spectral efficiency varying with SNR in different precoding schemes.

In Figure 4, the spectral efficiency of different precoding methods are simulated and compared under different numbers of RF chains. Through the data analysis of Table 3, we can see that when the number of RF chains equals to 10, the spectral efficiency of the algorithm proposed in this paper is 5% higher than that of the OMP-based joint

codebook algorithm [4] and 10% higher than that of the hybrid precoding algorithm with quantization [1]. When the number of RF chains equals 12, the spectral efficiency of the algorithm proposed is, at most, 6.3% higher than that of the OMP-based joint codebook algorithm [4] and 16.7% higher than that of the hybrid precoding algorithm with quantization [1]. As the number of RF chains increases, the spectral efficiency of the proposed codebook design algorithm gradually decreases, but it can still reach more than 83.9% of the optimal unconstrained precoder. All in all, the proposed codebook design algorithm generally performs better than the OMP-based joint codebook [4] algorithm and the hybrid precoding algorithm with quantization [1].

Table 3. Spectral efficiency of different algorithms with the number of RF chains.

Algorithm	Spectral Efficiency (bps/Hz)		
	$N_t\text{RF} = 6$	$N_t\text{RF} = 8$	$N_t\text{RF} = 12$
Hybrid Precoding with quantization [1]	13.85	13.12	11.86
OMP-based Codebook [4]	14.24	13.73	13.02
Proposed Codebook	15.10	14.43	13.84
Optimal unconstrained precoding	16.50	16.50	16.50

Figure 4. Spectral efficiency varying with the number of RF chains in different precoding schemes.

4.3. Comparison and Analysis of Computational Complexity

This paper mainly compares the performance of four algorithms. The proposed codebook design and OMP-based codebook [4] algorithms mainly include three steps. The first step is analog codebook design. The proposed codebook algorithm uses the CoSaMP algorithm to update iteratively, while the OMP-based codebook [4] algorithm uses the OMP algorithm to update iteratively. In the second step, both of them use dictionary learning algorithm for digital codebook design. In the third step, both of them use the Lloyd clustering algorithm for codebook space reduction. According to the conclusion of reference [21], the computational complexity of the CoSaMP algorithm is $\mathcal{O}(N_{\text{tr}}N_tN_s)$, while that of the OMP algorithm is $\mathcal{O}(N_{\text{tr}}N_tN_s^2)$. Therefore, we can conclude that the computational complexity of proposed codebook algorithm is lower than that of the OMP-based codebook [4]. The hybrid precoding with quantification [1] algorithm is also based on the OMP algorithm for analog codebook design, but their digital codebook design directly uses Grassmannian codebook, and does not use some clustering algorithms to

reduce codebook space, so hybrid precoding with quantification [1] algorithm is lower than the proposed codebook and OMP-based codebook [4] algorithm in terms of computational complexity. The optimal unconstrained precoding algorithm is directly based on SVD decomposition of channel **H** to get $\mathbf{F}_{opt}$. Compared with the first three algorithms, its computational complexity can be ignored, but this algorithm is not realizable in reality. It is only used for performance comparison and is regarded as an ideal benchmark.

5. Conclusions

In this paper, we focus on the design of the hybrid precoding codebook in FDD-based millimeter-wave massive MIMO systems. By improving the reconstruction algorithm and combining it with the dictionary learning algorithm, we propose a new hybrid precoding codebook construction algorithm to improve the spectral efficiency. Firstly, we reconstructed the digital precoding matrix according to the idea of CoSaMP algorithm, then combine the dictionary learning algorithm to update the analog codebook and digital precoding matrix, and then obtain the final digital precoding codebook through iterative optimization; finally, the qualified size digital precoding codebook is generated by using the clustering algorithm. The simulation results clearly demonstrate that the spectral efficiency of the proposed algorithm is higher than that of the OMP-based joint codebook [4] algorithm and the hybrid precoding with quantization [1] algorithm under low SNR and different numbers of RF chains. Particularly, when SNR is lower than 0 dB, the proposed algorithm is very close to the optimal unconstrained precoding algorithm because it carefully avoids the error accumulation effect by selecting multiple columns at each iteration. However, in the high-SNR region, compared with the other two algorithms, our algorithm has no great advantages and sometimes even equal spectral efficiency performance. This is because the algorithms in this paper have little error effect in iteration, which is caused by high SNR.

Author Contributions: Conceptualization, G.L. and K.Y.; methodology, G.L., H.D. (Honggui Deng) and G.L.; software, G.L. and K.Y.; formal analysis, G.L. and J.L.; investigation, G.L. and Z.Z.; data curation, G.L. and H.D. (Hu Dong); writing—original draft preparation, G.L. and K.Y.; writing—review and editing, G.L., K.Y. and H.D. (Honggui Deng). All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. El Ayach, O.; Rajagopal, S.; Abu-Surra, S.; Pi, Z.; Heath, R.W. Spatially Sparse Precoding in Millimeter Wave MIMO Systems. *IEEE Trans. Wirel. Commun.* **2014**, *13*, 1499–1513. [CrossRef]
2. Alkhateeb, A.; Mo, J.; Gonzalez-Prelcic, N.; Heath, R.W. MIMO Precoding and Combining Solutions for Millimeter-Wave Systems. *IEEE Commun. Mag.* **2014**, *52*, 122–131. [CrossRef]
3. Kolawole, O.Y.; Biswas, S.; Singh, K.; Ratnarajah, T. Transceiver Design for Energy-Efficiency Maximization in mmWave MIMO IoT Networks. *IEEE Trans. Green Commun. Netw.* **2020**, *4*, 109–123. [CrossRef]
4. Liu, X.; Zou, W.; Chen, S. Joint Design of Analog and Digital Codebooks for Hybrid Precoding in Millimeter Wave Massive MIMO Systems. *IEEE Access* **2018**, *6*, 69818–69825. [CrossRef]
5. Li, N.; Wei, Z.; Yang, H.; Zhang, X.; Yang, D. Hybrid Precoding for mmWave Massive MIMO Systems With Partially Connected Structure. *IEEE Access* **2017**, *5*, 15142–15151. [CrossRef]
6. Hu, C.; Zhang, J. Hybrid Precoding Design for Adaptive Subconnected Structures in Millimeter-Wave MIMO Systems. *IEEE Syst. J.* **2019**, *13*, 137–146. [CrossRef]
7. Zhang, D.; Wang, Y.; Li, X.; Xiang, W. Hybridly Connected Structure for Hybrid Beamforming in mmWave Massive MIMO Systems. *IEEE Trans. Commun.* **2018**, *66*, 662–674. [CrossRef]

8. Li, X.; Alkhateeb, A. Deep Learning for Direct Hybrid Precoding in Millimeter Wave Massive MIMO Systems. In Proceedings of the 2019 53rd Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 3–6 November 2019; pp. 800–805. [CrossRef]

9. Chen, J.; Wen, C.; Jin, S.; Wong, K. A Low Complexity Pilot Scheduling Algorithm for Massive MIMO. *IEEE Wirel. Commun. Lett.* **2017**, *6*, 18–21. [CrossRef]

10. Chen, C.H.; Tsai, C.R.; Liu, Y.H.; Hung, W.L.; Wu, A.Y. Compressive Sensing (CS) Assisted Low-Complexity Beamspace Hybrid Precoding for Millimeter-Wave MIMO Systems. *IEEE Trans. Signal Process.* **2017**, *65*, 1412–1424. [CrossRef]

11. Gao, X.; Dai, L.; Han, S.; Chih-Lin, I.; Heath, R.W. Energy-Efficient Hybrid Analog and Digital Precoding for MmWave MIMO Systems With Large Antenna Arrays. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 998–1009. [CrossRef]

12. Rusu, C.; Mendez-Rial, R.; Gonzalez-Prelcicy, N.; Heath, R.W. Low complexity hybrid sparse precoding and combining in millimeter wave MIMO systems. In Proceedings of the IEEE International Conference on Communications, Kuala Lumpur, Malaysia, 23–27 May 2016; pp. 1340–1345. [CrossRef]

13. Seleem, H.; Sulyman, A.; Alsanie, A. Hybrid Precoding-Beamforming Design with Hadamard RF Codebook for mmWave Large-Scale MIMO Systems. *IEEE Access* **2017**, *5*, 6813–6823.

14. Bai, T.; Alkhateeb, A.; Heath, R.W. Coverage and capacity of millimeter-wave cellular networks. *IEEE Commun. Mag.* **2014**, *52*, 70–77.

15. Pi, Z.; Khan, F. An introduction to millimeter-wave mobile broadband systems. *IEEE Commun. Mag.* **2011**, *49*, 101–107.

16. Alkhateeb, A.; El Ayach, O.; Leus, G.; Heath, R.W. Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems. *IEEE J. Sel. Top. Signal Process.* **2014**, *8*, 831–846.

17. Raghavan, V.; Sayeed, A.M. Sublinear Capacity Scaling Laws for Sparse MIMO Channels. *IEEE Trans. Inf. Theory* **2010**, *57*, 345–364. [CrossRef]

18. Palomar, D.P.; Cioffi, J.M.; Lagunas, M.A. Joint Tx-Rx Beamforming Design for Multicarrier MIMO Channels: A Unified Framework for Convex Optimization. *IEEE Trans. Signal Process.* **2003**, *51*, 2381–2401. [CrossRef]

19. Palomar, D.P.; Chiang, M. A Tutorial on Decomposition Methods for Network Utility Maximization (Tutorial Paper). *IEEE J. Sel. Areas Commun.* **2006**, *24*, 1439–1451. [CrossRef]

20. Love, D.J.; Heath, R.W. Limited feedback unitary precoding for spatial multiplexing systems. *IEEE Trans. Inf. Theory* **2005**, *51*, 2967–2976.

21. Needell, D.; Tropp, J. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmon. Anal.* **2009**, *26*, 301–321.