Low Complexity Hybrid Precoding for Millimeter Wave MIMO Systems

Mengying Jiang, Jiayan Zhang, and Honglin Zhao(✉)

Harbin Institute of Technology, Harbin, China
hlzhao@hit.edu.cn

Abstract. Massive multiple input multiple output (MIMO) technology and millimeter wave (mmWave) system, as the key technologies of the new generation of mobile communications, can effectively increase channel capacity and relieve spectrum resources. Because the mmWave has a short wavelength, the transceiver can be composed of a large antenna array to reduce severe signals attenuation. Furthermore, the use of hybrid precoding technology can improve system performance and reduce system hardware complexity. The classic hybrid precoding algorithm that based on simultaneous orthogonal matching pursuit (SOMP) requires matrix inversion, which leads to high complexity, and its performance depends on the accuracy of channel estimation. In this paper, by modeling the mmWave MIMO system, we compare three improved algorithms, which are orthogonality based matching pursuit algorithm (OBMP), matrix-inversion-bypass simultaneous orthogonal matching pursuit algorithm (MIB-SOMP) and residual matrix-singular value decomposition algorithm (RM-SVD). We analyze the performance of the algorithms, such as complexity, spectrum efficiency, bit error rate, as well as the advantages and disadvantages of the algorithms.

Keywords: Massive MIMO · Millimeter wave · Hybrid precoding · Low complexity

1 Introduction

The millimeter wave (mmWave) multiple input multiple output (MIMO) system is a promising technology to achieve data rates above 10Gbit/s [1]. The most prominent feature of mmWave system is their huge bandwidth, which can support high data rates communications. However, compared with the frequency bands of existing cellular communications, mmWave signals have severe path loss. But because of the shorter wavelength of mmWave, more antennas can be integrated on the same physical size to improve antenna gain [2].

An effective solution to achieve high link quality and support high data rates is to transmit multiple data streams through precoding technology [3]. In 4G mobile communications, traditional digital precoding technology has been widely used [4]. However, it is not practical to use traditional digital precoding on a large-scale antenna array, because digital precoding requires a separate radio frequency (RF) chain for each antenna, which will greatly increase the cost and power consumption of the system.

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2022
Published by Springer Nature Switzerland AG 2022. All Rights Reserved
Q. Guo et al. (Eds.): WiSATS 2021, LNICST 410, pp. 418–431, 2022.
https://doi.org/10.1007/978-3-030-93398-2_40

Generally, hybrid precoding technology is used in mmWave MIMO system, which combines baseband precoding with RF precoding to solve the problems of high hardware cost and power consumption.

The hybrid precoding design can be solved by simultaneous orthogonal matching pursuit (SOMP) [5], an algorithm requiring matrix inversion and assuming that the channel state information (CSI) is perfectly known. Hence, several approaches are proposed to simplify the matrix inversion computation and avoid the dependence on channel estimation. Based on the singular value decomposition (SVD), a hybrid precoding algorithm based on the SVD of the residual matrix is proposed [6], which reduces the dependence on channel estimation. And in [7], the DFT codebook is used as a candidate matrix for its orthogonality, which can avoid the iterations in matching pursuit. In addition, an approach is proposed to simplify the matrix inversion computation by applying Schur-Banachiewicz blockwise inversion [8].

On this basis, we compare the three improved algorithms, quantitatively analyze the complexity and the performance of these algorithms, as well as the existing problems.

This paper is organized as follows. The system model, channel model and the classic hybrid precoding algorithm are introduced in Sect. 2. The RM-SVD, OBMP and MIB-SOMP algorithms are presented in Sect. 3. The simulation analysis and performance comparison of the three improved algorithms are presented in Sect. 4. Finally, the conclusions about the performance analysis are presented in Sect. 5.

2 System Model

2.1 System Model

Consider the single-user mmWave MIMO system shown in Fig. 1 [9].

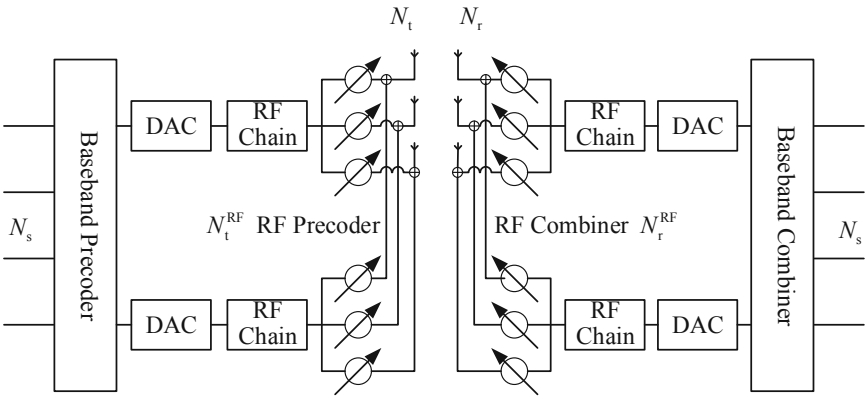


Fig. 1. MmWave MIMO hybrid precoding system block diagram

The base station equipped with N_t antennas transmit N_s data streams to the user equipment equipped with N_r antennas. There are N_t^{RF} and N_r^{RF} RF chains at the transmitter and the receiver to enable multi-stream transmission such that $N_s \le N_t^{RF} \le N_t$, $N_s \le N_r^{RF} \le N_r$. In the system, the signal processed by the baseband precoder $\mathbf{F}_{BB} \in \mathbb{C}^{N_t^{RF} \times N_s}$ and RF precoder $\mathbf{F}_{RF} \in \mathbb{C}^{N_t \times N_t^{RF}}$ is transmitted to the channel $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$. The transmitted signal at the base station is:

$$\mathbf{x} = \mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{s} \tag{1}$$

where $\mathbf{s} = [s_1, s_2, \ldots, s_{N_s}]^T$ denotes the data stream such that $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = \frac{1}{N_s}\mathbf{I}_{N_s}$. The total transmit power constraint satisfies $\|\mathbf{F}_{RF}\mathbf{F}_{BB}\|_F^2 = N_s$. The received signal at the user equipment is:

$$\mathbf{y} = \sqrt{\rho}\mathbf{H}\mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{s} + \mathbf{n} \tag{2}$$

where ρ represents the average received power, $\mathbf{n} \sim \mathcal{CN}\left(0, \sigma_n^2\mathbf{I}_{N_r}\right)$ is the noise, ρ/σ_n^2 is the received SNR, and $\mathbf{H}$ is the channel matrix such that $\mathbb{E}\left[\|\mathbf{H}\|_F^2\right] = N_t N_r$. After being processed by the RF combiner $\mathbf{W}_{RF} \in \mathbb{C}^{N_r \times N_r^{RF}}$ and baseband combiner $\mathbf{W}_{BB} \in \mathbb{C}^{N_r^{RF} \times N_s}$, the received signal is:

$$\hat{\mathbf{y}} = \sqrt{\rho}\mathbf{W}_{BB}^H\mathbf{W}_{RF}^H\mathbf{H}\mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{s} + \mathbf{W}_{BB}^H\mathbf{W}_{RF}^H\mathbf{n} \tag{3}$$

Since $\mathbf{F}_{RF}$ and $\mathbf{W}_{RF}$ is implemented using analog phase shifters, its elements are constrained to satisfy $\left|\mathbf{F}_{RF}^{(i,j)}\right| = 1/\sqrt{N_t}$ and $\left|\mathbf{W}_{RF}^{(i,j)}\right| = 1/\sqrt{N_r}$.

2.2 Millimeter Wave Channel Model

Considering that the mmWave has the characteristics of high path loss and line-of-sight transmission, the narrowband scattering cluster channel model is used to model the smmWave channel [10]. Assuming that the mmWave channel contains N_{cl} scattering cluster, each cluster contains N_{ray} propagation path, the system channel can be described as:

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{N_{cl} N_{ray}}} \sum_{i,l} \alpha_{i,l}\mathbf{a}_r\left(\phi_{il}^r, \theta_{il}^r\right)\mathbf{a}_t\left(\phi_{il}^t, \theta_{il}^t\right)^H \tag{4}$$

where $\alpha_{i,l} \sim \mathcal{CN}\left(0, \sigma_{\alpha,i}^2\right)$ is the complex path gain of the l^{th} ray in the i^{th} scattering cluster, and $\sigma_{\alpha,i}^2$ represents the average power of the i^{th} cluster. The average cluster powers are such that $\sum_{i=1}^{N_{cl}} \sigma_{\alpha,i}^2 = \gamma$ where γ satisfies $\mathbb{E}\left[\|\mathbf{H}\|_F^2\right] = N_t N_r$. ϕ_{il}^r and θ_{il}^r represent the azimuth and elevation angles of the arrival angle. In the same way, ϕ_{il}^t and θ_{il}^t

represent the azimuth and elevation angles of the departure angle. $\mathbf{a}_{\mathrm{r}}\left(\phi_{il}^{\mathrm{r}}, \theta_{il}^{\mathrm{r}}\right)$ and $\mathbf{a}_{\mathrm{t}}\left(\phi_{il}^{\mathrm{t}}, \theta_{il}^{\mathrm{t}}\right)$ are the antenna array response vector at the receiver and the transmitter.

For the convenience of description, formula (4) can be written as:

$$\mathbf{H} = \mathbf{A}_{\mathrm{r}}\mathbf{H}_{\mathrm{a}}\mathbf{A}_{\mathrm{t}}^{H} \tag{5}$$

where $\mathbf{H}_{\mathrm{a}}=\sqrt{N_{\mathrm{t}}N_{\mathrm{r}}/N_{\mathrm{cl}}N_{\mathrm{ray}}}\,\mathrm{diag}\left(\alpha_{1,1}, \alpha_{1,2}, \cdots, \alpha_{cl,ray}\right)$ represents the path gain of all propagation paths in the channel. $\mathbf{A}_{\mathrm{t}}$ and $\mathbf{A}_{\mathrm{r}}$ represent the transmit candidate matrix and receive candidate matrix.

In this paper, we adopt uniform linear arrays and then the array response vector is:

$$\mathbf{a}(\phi) = \sqrt{\frac{1}{N}}\left[1, e^{-jkd\sin\phi}, e^{-j2kd\sin\phi}, \ldots, e^{-j(N-1)kd\sin\phi}\right]^{T} \tag{6}$$

where N is the number of antennas, d is the antenna spacing, and $k = 2\pi/\lambda$.

2.3 Hybrid Precoder Designs

The target of hybrid precoding is to maximize the spectral efficiency over all possible solutions of $(\mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{BB}}, \mathbf{W}_{\mathrm{RF}}, \mathbf{W}_{\mathrm{BB}})$, which is written as:

$$R = \log_2\left|\mathbf{I}_{N_{\mathrm{s}}} + \frac{\rho}{N_{\mathrm{s}}}\mathbf{R}_{\mathrm{n}}^{-1}\mathbf{W}_{\mathrm{BB}}^{H}\mathbf{W}_{\mathrm{RF}}^{H}\mathbf{H}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\mathbf{F}_{\mathrm{BB}}^{H}\mathbf{F}_{\mathrm{RF}}^{H}\mathbf{H}^{H}\mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}\right| \tag{7}$$

where $\mathbf{R}_{\mathrm{n}} = \sigma_{\mathrm{n}}^2\mathbf{W}_{\mathrm{BB}}^{H}\mathbf{W}_{\mathrm{RF}}^{H}\mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}$ represents the noise covariance matrix processed by the user equipment. According to mathematical derivation, the joint optimization problem can be decoupled. Then we only consider the design of the hybrid precoder $\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}$ and assume the receiver can decode perfectly.

In general, the precoder design problem can be rewritten as:

$$\begin{aligned}\left(\mathbf{F}_{\mathrm{RF}}^{\mathrm{opt}}, \mathbf{F}_{\mathrm{BB}}^{\mathrm{opt}}\right) &= \arg\min_{\mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{BB}}}\left\|\mathbf{F}_{\mathrm{opt}} - \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\right\|_{F} \\ \text{s.t.} \quad & \mathbf{F}_{\mathrm{RF}} \in \varphi, \ \ \left\|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\right\|_{F}^{2} = N_{\mathrm{s}}\end{aligned} \tag{8}$$

where $\mathbf{F}_{\mathrm{opt}}$ represents the optimal full-digital precoder, which consists of right singular vectors associated with the largest N_{s} eigenvalues of $\mathbf{H}$.

Similarly, considering the fixed hybrid precoder $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{F}_{\mathrm{BB}}$, we adopt the minimum mean square error (MMSE) criterion to design the combiner $\mathbf{W}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{BB}}$. According to derivation, the design problem of the combiner can be equivalent to [5]:

$$\begin{aligned}\left(\mathbf{W}_{\mathrm{RF}}^{\mathrm{opt}}, \mathbf{W}_{\mathrm{BB}}^{\mathrm{opt}}\right) &= \underset{\mathbf{W}_{\mathrm{RF}}, \mathbf{W}_{\mathrm{BB}}}{\mathrm{argmin}}\left\|\mathbb{E}\left[yy^{H}\right]^{1/2}\left(\mathbf{W}_{\mathrm{MMSE}} - \mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}\right)\right\|_{F} \\ \text{s.t.} \quad & \mathbf{W}_{\mathrm{RF}} \in \varphi\end{aligned} \tag{9}$$

Therefore, the design of the MMSE combiner is similar to the design of the hybrid precoder. The differences are that there is a weighting of the received signal power $\mathbb{E}[yy^H]$ in the optimization objective function and there is no power limitation for the combiner. In this paper, we adopt the MMSE criterion to convert the precoding algorithm to the receiver to solve the design problem of the combiner.

The hybrid precoding algorithm based on SOMP is shown in Table 1.

Table 1. The Hybrid precoding algorithm based on SOMP

Algorithm 1: SOMP
Input $\mathbf{F}_{\text{opt}}$ and $\mathbf{A}_{\text{t}}$
Output $\mathbf{F}_{\text{RF}}$ and $\mathbf{F}_{\text{BB}}$
1: $\mathbf{F}_{\text{RF}} = [\ ]$
2: $\mathbf{F}_{\text{res}} = \mathbf{F}_{\text{opt}}$
3: for $i = 1 : N_{\text{t}}^{\text{RF}}$ do
4: $\boldsymbol{\Psi}_i = \mathbf{A}_{\text{t}}^H \mathbf{F}_{\text{res}}$
5: $k = \arg\max_{l=1,\ldots,N_{\text{cl}}N_{\text{ray}}} \left(\boldsymbol{\Psi}_i \boldsymbol{\Psi}_i^H\right)_{l,l}$
6: $\mathbf{F}_{\text{RF}} = \left[\mathbf{F}_{\text{RF}} \middle| \mathbf{A}_{\text{t}}^{(k)}\right]$
7: $\mathbf{F}_{\text{BB}} = \left(\mathbf{F}_{\text{RF}}^H \mathbf{F}_{\text{RF}}\right)^{-1} \mathbf{F}_{\text{RF}}^H \mathbf{F}_{\text{opt}}$
8: $\mathbf{F}_{\text{res}} = \dfrac{\mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}}{\left\|\mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\right\|_F}$
9: end for
10: $\mathbf{F}_{\text{BB}} = \sqrt{N_{\text{s}}} \dfrac{\mathbf{F}_{\text{BB}}}{\left\|\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\right\|_F}$

The algorithm mainly has two issues: 1) High complexity of matrix inversion. SOMP algorithm requires matrix inversion for updating the baseband precoder with least squares method. When the number of selected base vectors increases, the dimension of matrix inversion will also increase, and matrix inversion is complicated for hardware implementation, which may cause more long calculation delay and higher power consumption. 2) High dependence on the channel estimation. In this algorithm, the RF precoder is the premise of the baseband precoder. And because of the constant modulus constraints and power control, the design of RF precoder is the main factor

affecting system performance. This algorithm uses the antenna array response vector as a candidate matrix for RF precoder, and the system performance will be affected by the accuracy of the channel estimation.

3 Hybrid Precoding Algorithm

3.1 RM-SVD

The RM-SVD algorithm is mainly composed of two parts: the design of the initial RF precoding matrix and the update of the initial precoding matrix. The initial RF precoding matrix is constructed by performing SVD on the optimal full-digital precoding matrix $\mathbf{F}_{\text{opt}}$, as shown in Eq. (10):

$$\mathbf{F}_{\text{opt}} = \mathbf{SVD}^H \tag{10}$$

where $\mathbf{SV} \in \mathbb{C}^{N_t \times N_s}$. In order to determine the initial RF precoding matrix, a $N_t \times \left(N_t^{RF} - N_s\right)$ dimensional matrix is constructed so that the phase of its elements obeys a uniform random distribution on $[0, 2\pi)$. At the same time, it is forced to limit its amplitude to $1/\sqrt{N_t}$ for meeting the constant modulus constraints of $\mathbf{F}_{\text{RF}}$. Therefore, the Eq. (10) has the following equivalent form:

$$\mathbf{F}_{\text{opt}} = [\mathbf{SVF}_{\text{R}}] \begin{bmatrix} \mathbf{D}^H \\ \mathbf{0} \end{bmatrix} \tag{11}$$

where $\mathbf{0}$ represents an all-zero matrix with $\left(N_t^{RF} - N_s\right) \times N_s$ dimensions. Under unrestricted conditions, according to the SVD of $\mathbf{F}_{\text{opt}}$, a global optimal solution can be obtained:

$$\mathbf{F}_{\text{RF}}^* = [\mathbf{SVF}_{\text{R}}],\ \mathbf{F}_{\text{BB}}^* = \begin{bmatrix} \mathbf{D}^H \\ \mathbf{0} \end{bmatrix} \tag{12}$$

But the $\mathbf{SV}$ does not meet the constant modulus constraints. Therefore, the phases of all elements of the $\mathbf{SV}$ are preserved, and the amplitude of all elements is forced to be $1/\sqrt{N_t}$. The processed $\mathbf{F}_{\text{RF}}^*$ is used as the initial RF precoding matrix.

The hybrid precoding algorithm based on RM-SVD is shown in Table 2.

Table 2. The Hybrid precoding algorithm based on RM-SVD

Algorithm 2: RM-SVD
Input $\mathbf{F}_{\text{opt}}$ and initial $\mathbf{F}_{\text{RF}}$
Output $\mathbf{F}_{\text{RF}}$ and $\mathbf{F}_{\text{BB}}$
1: for $i = 1 : N_{\text{t}}^{\text{RF}}$ do
2: $\mathbf{F}_{\text{RF}}(:,1) = [\ ]$
3: $\mathbf{F}_{\text{BB}} = \left(\mathbf{F}_{\text{RF}}^{H}\mathbf{F}_{\text{RF}}\right)^{-1}\mathbf{F}_{\text{RF}}^{H}\mathbf{F}_{\text{opt}}$
4: $\mathbf{F}_{\text{res}} = \mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}$
5: $\mathbf{F}_{\text{res}} = \mathbf{USV}^{H}$
6: $\mathbf{n} = 1/\sqrt{N_{\text{t}}} \cdot \mathbf{e}^{j\arg(\mathbf{U}(:,1))}$
7: $\mathbf{F}_{\text{RF}} = [\mathbf{F}_{\text{RF}}\ \mathbf{n}]$
8: end for
9: $\mathbf{F}_{\text{BB}} = \left(\mathbf{F}_{\text{RF}}^{H}\mathbf{F}_{\text{RF}}\right)^{-1}\mathbf{F}_{\text{RF}}^{H}\mathbf{F}_{\text{opt}}$
10: $\mathbf{F}_{\text{BB}} = \sqrt{N_{\text{s}}}\dfrac{\mathbf{F}_{\text{BB}}}{\left\|\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\right\|_{F}}$

Compared with the SOMP algorithm, this algorithm does not require a candidate matrix, which reduces the dependence on channel estimation. Furthermore, through the SVD of the residual matrix, the information of the residual matrix can be better used, so that the hybrid precoder $\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}$ is closer to $\mathbf{F}_{\text{opt}}$. However, due to matrix inversion and SVD, the algorithm complexity is relatively high.

3.2 OBMP

To avoid the issues of the SOMP algorithm, the candidate matrix must satisfy the orthogonality of the column vectors. And considering the sparse characteristics of the mmWave channel, the DFT codebook is used as the candidate matrix, denoted as $\mathbf{A}_{\text{t,DFT}}$

$$\mathbf{A}_{\text{t,DFT}}(:,k) = \frac{1}{\sqrt{N_{\text{t}}}}\left[1, e^{\frac{-j2\pi(k-1)}{N_{\text{t}}}}, \ldots, e^{\frac{-j2\pi(k-1)(N_{\text{t}}-1)}{N_{\text{t}}}}\right]^{T} \tag{13}$$

where $k = 1, \ldots, N_{\text{t}}$. It is well known that the DFT codebook is the basis for the space that is spanned by array response vectors.

Based on the orthogonality of the DFT codebook, as shown in Fig. 2, by calculating the correlation matrix of $\mathbf{F}_{\text{opt}}$ and $\mathbf{A}_{\text{t,DFT}}$, the power contribution distributed in N_{t}

beamforming directions can be calculated in parallel. The correlation matrix can be expressed as:

$$\mathbf{\Psi}_0 = \mathbf{A}_{\mathrm{t,DFT}}^{H}\mathbf{F}_{\mathrm{opt}} \tag{14}$$

The power contribution of each beamforming direction can be obtained according to formula (18):

$$\boldsymbol{\beta} = \mathrm{diag}\left(\mathbf{\Psi}_0\mathbf{\Psi}_0^{H}\right) \tag{15}$$

where the i^{th} element of $\boldsymbol{\beta} \in \mathbb{C}^{N_t \times 1}$ represents the contribution of the i^{th} beamforming direction in the DFT codebook. Because of the orthogonality of the DFT beamforming vectors, the N_t^{RF} column vectors can be selected in parallel in the DFT codebook without iteration to form $\mathbf{F}_{\mathrm{RF}}$:

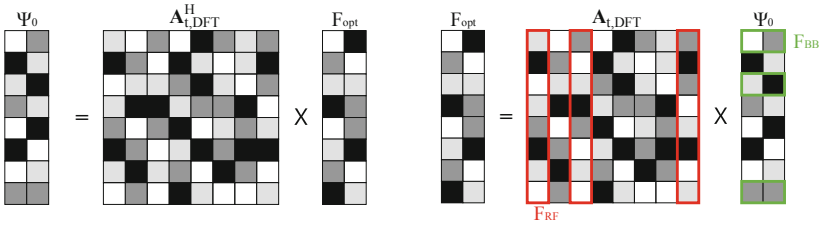
$$\mathbf{F}_{\mathrm{RF}} = \mathbf{A}_{\mathrm{t,DFT}}(:, \mathcal{V}) \tag{16}$$

where $\mathcal{V}$ is the index set of the largest N_t^{RF} elements selected in $\boldsymbol{\beta}$. Since $\mathbf{F}_{\mathrm{RF}} = \mathbf{A}_{\mathrm{t,DFT}}(:, \mathcal{V})$ is composed of orthogonal column vectors in the DFT codebook, the inverse matrix of $\mathbf{F}_{\mathrm{RF}}^{H}\mathbf{F}_{\mathrm{RF}}$ is the identity matrix. Then $\mathbf{F}_{\mathrm{BB}}$ can be expressed as:

$$\begin{aligned}\mathbf{F}_{\mathrm{BB}} &= \left(\mathbf{A}_{\mathrm{t,DFT}}(:, \mathcal{V})^{H}\mathbf{A}_{\mathrm{t,DFT}}(:, \mathcal{V})\right)^{-1}\mathbf{A}_{\mathrm{t,DFT}}(:, \mathcal{V})^{H}\mathbf{F}_{\mathrm{opt}} \\ &= \mathbf{A}_{\mathrm{t,DFT}}(:, \mathcal{V})^{H}\mathbf{F}_{\mathrm{opt}} = \mathbf{\Psi}_0(\mathcal{V}, :)\end{aligned} \tag{17}$$

As shown in Fig. 2, the least squares method can be simplified as selecting rows in $\mathbf{\Psi}_0$ with $\mathcal{V}$ as the index set.

The hybrid precoding algorithm based on OBMP is shown in Table 3.



(a) Correlation matrix (b) Parallel index selection

Fig. 2. The explanation of OBMP algorithm

Table 3. The Hybrid precoding algorithm based on OBMP

Algorithm 3: OBMP
Input $\mathbf{F}_{\text{opt}}$ and $\mathbf{A}_{\text{t,DFT}}$
Output $\mathbf{F}_{\text{RF}}$ and $\mathbf{F}_{\text{BB}}$
1: $\mathbf{\Psi}_0 = \mathbf{A}_{\text{t,DFT}}^H \mathbf{F}_{\text{opt}}$
2: $\boldsymbol{\beta} = \text{diag}\left(\mathbf{\Psi}_0 \mathbf{\Psi}_0^H\right)$
3: $\mathcal{V}_0 = [\ ]$
4: for $i = 1 : N_{\text{t}}^{\text{RF}}$ do
5: $k = \arg\max_{l=1,\ldots,N_{\text{t}}} \boldsymbol{\beta}(l)$
6: $\mathcal{V}_i = [\mathcal{V}_{i-1}\ k]$
7: $\boldsymbol{\beta}(k) = 0$
8: end for
9: $\mathbf{F}_{\text{RF}} = \mathbf{A}_{\text{t,DFT}}(:, \mathcal{V})$
10: $\mathbf{F}_{\text{BB}} = \mathbf{\Psi}_0(\mathcal{V}, :)$
11: $\mathbf{F}_{\text{BB}} = \sqrt{N_{\text{s}}} \dfrac{\mathbf{F}_{\text{BB}}}{\left\|\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}\right\|_F}$

3.3 MIB-SOMP

In algorithm MIB-SOMP, Schur-Banachiewicz blockwise inversion is used to avoid complex matrix inversion. The algorithm divides a high-dimensional matrix into a low-dimensional matrix, and uses the result of the previous iteration to perform the inversion operation of the block matrix in the current iteration, which reduce the complexity of the SOMP algorithm.

In this algorithm, we can use auxiliary variables $\mathbf{A}, V, \mathbf{M}$ and the calculation result of the previous iteration to update $\mathbf{G}_{\mathcal{I}_i,\mathcal{I}_i}^{-1}$, $\mathbf{F}_{\text{BB}_i}$ and $\mathbf{\Psi}_i$. And the matrix inversion is converted to matrix multiplication, which improves the hardware realizability of the algorithm.

According to theoretical derivation, the hybrid precoding algorithm based on MIB-SOMP is shown in Table 4.

Table 4. The Hybrid precoding algorithm based on MIB-SOMP

Algorithm 4: MIB-SOMP

Input $\mathbf{F}_{\text{opt}}$ and $\mathbf{A}_{\text{t}}$

Output $\mathbf{F}_{\text{RF}}$ and $\mathbf{F}_{\text{BB}}$

1: $\mathbf{G} = \mathbf{A}_{\text{t}}^{H}\mathbf{A}_{\text{t}}\left(\mathbf{G}_{\mathcal{I},\mathcal{J}} = \mathbf{A}_{\text{t}_{\mathcal{I}}}^{H}\mathbf{A}_{\text{t}_{\mathcal{J}}}\right)$

2: $\mathbf{G}_{\mathcal{I}_0,\mathcal{I}_0}^{-1} = \mathbf{F}_{\text{BB}_0} = [\ ],\ \boldsymbol{\Psi}_0 = \mathbf{A}_{\text{t}}^{H}\mathbf{F}_{\text{opt}}$

3: $\mathcal{I}_0 = [\ ],\ \overline{\mathcal{I}}_0 = \left\{i,\ 1 \le i \le N_{\text{cl}}N_{\text{ray}}\right\}$

4: for $i = 1 : N_{\text{t}}^{\text{RF}}$ do

5: $k = \arg\max_l \left(\boldsymbol{\Psi}_{i-1}\boldsymbol{\Psi}_{i-1}^{H}\right)_{l,l}$

6: $\mathbf{A} = \mathbf{G}_{k,\mathcal{I}_{i-1}}\mathbf{G}_{\mathcal{I}_{i-1},\mathcal{I}_{i-1}}^{-1}$

7: $V = 1/\left(\mathbf{G}_{k,k} - \mathbf{A}\mathbf{G}_{\mathcal{I}_{i-1},k}\right)$

8: $\mathbf{M} = \mathbf{A}\boldsymbol{\Psi}_0\left(\mathcal{I}_{i-1},:\right) - \boldsymbol{\Psi}_0\left(k,:\right)$

9: $\mathcal{I}_i = \left[\mathcal{I}_{i-1}\ k\right],\ \overline{\mathcal{I}}_i = \overline{\mathcal{I}}_{i-1} - \{k\}$

10: $\mathbf{G}_{\mathcal{I}_i,\mathcal{I}_i}^{-1} = \begin{bmatrix} \mathbf{G}_{\mathcal{I}_{i-1},\mathcal{I}_{i-1}}^{-1} + V\mathbf{A}^{H}\mathbf{A} & -V\mathbf{A}^{H} \\ -V\mathbf{A} & V \end{bmatrix}$

11: $\mathbf{F}_{\text{BB}_i} = \begin{bmatrix} \mathbf{F}_{\text{BB}_{i-1}} + V\mathbf{A}^{H}\mathbf{M} \\ -V\mathbf{M} \end{bmatrix}$

12: $\boldsymbol{\Psi}_i = \boldsymbol{\Psi}_{i-1}\left(\overline{\mathcal{I}}_i,:\right) - \mathbf{G}_{\overline{\mathcal{I}}_i,\mathcal{I}_i}\begin{bmatrix} V\mathbf{A}^{H}\mathbf{M} \\ -V\mathbf{M} \end{bmatrix}$

13: end for

14: $\mathbf{F}_{\text{RF}} = \mathbf{A}_{\text{t}_{\mathcal{I}_{N_{\text{t}}^{\text{RF}}}}}$

15: $\mathbf{F}_{\text{BB}} = \sqrt{N_{\text{s}}}\,\dfrac{\mathbf{F}_{\text{BB}}}{\left\|\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\right\|_F}$

4 Performance Analysis

4.1 Complexity Analysis

Compared with addition, multiplication calculation is the main factor of the complexity of the algorithm. Therefore, in this paper we only consider the number of complex multiplications in each iteration.

The number of complex multiplications during the i^{th} iteration of the hybrid precoding algorithm of SOMP and RM-SVD is shown in Table 5 and Table 6.

Table 5. Number of complex multiplications in i^{th} the iteration of SOMP

Computation	SOMP
$\mathbf{\Psi}_i$	$L \times N_\mathrm{t} \times N_\mathrm{s}$
$\mathbf{\Psi}_i \mathbf{\Psi}_i^H$	$L \times N_\mathrm{s} \times L$
$\mathbf{F}_\mathrm{RF}\mathbf{F}_\mathrm{BB}$	$N_\mathrm{t} \times N_\mathrm{t}^\mathrm{RF} \times N_\mathrm{s}$
$\left\|\mathbf{F}_\mathrm{opt} - \mathbf{F}_\mathrm{RF}\mathbf{F}_\mathrm{BB}\right\|_F$	$N_\mathrm{t} \times N_\mathrm{s}$

Table 6. Number of complex multiplications in i^{th} the iteration of RM-SVD

Computation	RM-SVD
$\mathbf{F}_\mathrm{RF}\mathbf{F}_\mathrm{BB}$	$N_\mathrm{t} \times N_\mathrm{t}^\mathrm{RF} \times N_\mathrm{s}$
$\mathbf{USV}^H$	$N_\mathrm{t} \times N_\mathrm{s} \times N_\mathrm{t}$
$\mathbf{n} = 1/\sqrt{N_\mathrm{t}} \cdot \mathbf{e}^{j\arg(\mathbf{U}(:,1))}$	N_t

where L is the number of basis vectors in the candidate matrix, N_t is the number of antennas at the transmitter, N_s is the number of data streams, and N_t^RF is the number of RF chains at the transmitter. The complexity of the SOMP algorithm is mainly concentrated in the calculation of the correlation matrix and the residual matrix. Furthermore, the complexity of the RM-SVD algorithm is mainly concentrated in the SVD of the residual matrix. Due to the sparsity of mmWave channels, N_t is much larger than L in mmWave MIMO systems, so the RM-SVD algorithm requires more complex multiplications.

The number of complex multiplications during the i^{th} iteration of the hybrid precoding algorithm of OBMP and MIB-SOMP is shown in Table 7 and Table 8.

Table 7. Number of complex multiplications in i^{th} the iteration of OBMP

Computation	OBMP
$\mathbf{\Psi}_0$	$N \times N \times N_\mathrm{s}$
$\boldsymbol{\beta}$	$N \times N_\mathrm{s}/2$

Table 8. Number of complex multiplications in i^{th} the iteration of MIB-SOMP

Computation	MIB-SOMP	Computation	MIB-SOMP
$\mathbf{\Psi}_0$	$L \times N \times N_\mathrm{s}$	$\mathbf{M}$	$(i-1) \times N_\mathrm{s}$
$\mathbf{G}$	$L \times N \times L$	$\mathbf{A}^H\mathbf{V}$	$i-1$
$\left(\mathbf{\Psi}_{i-1}\mathbf{\Psi}_{i-1}^H\right)_{l,l}$	$[L-(i-1)] \times N_\mathrm{s}/2$	$\left(\mathbf{A}^H\mathbf{V}\right)\mathbf{A}$	$(i-1) \times (i-1)$
$\mathbf{\Psi}_i$	$(L-i) \times i \times N_\mathrm{s}$	$\left(\mathbf{A}^H\mathbf{V}\right)\mathbf{M}$	$(i-1) \times N_\mathrm{s}$
$\mathbf{A}$	$(i-1) \times (i-1)$	$\mathbf{VM}$	N_s
$\mathbf{V}$	$i-1$	$\mathbf{G}_i^{-1}$	0

The OBMP algorithm uses the DFT codebook with orthogonal column vectors as the candidate matrix and selects N_t^{RF} column vectors with the largest power contribution in parallel. In addition, the algorithm converts the least square method into a simple indexing process so that matrix inversion is not required, which greatly reduces the complexity of the algorithm.

The MIB-SOMP algorithm applies Schur-Banachiewicz blockwise inversion and achieves hybrid precoding through iterative inversion. This procedure greatly reduces the computation complexity and enable efficient hardware implementation.

4.2 Spectral Efficiency

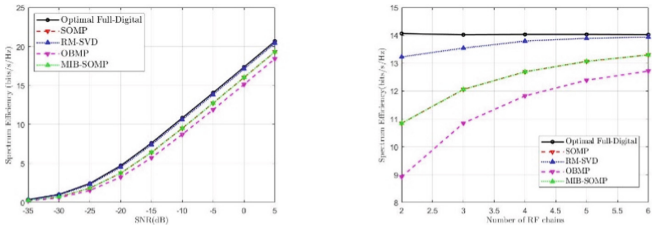
The simulation parameters in the performance analysis are shown in Table 9.

Table 9. The simulation parameters

Parameters	Values
$N_t \times N_r$	128×32
d	0.5λ
Modulation scheme	QPSK
N_{cl}	5
N_{ray}	10

In Fig. 3(a), we compare the spectral efficiency of all hybrid precoding algorithms. This simulation sets $N_s = 2$ and $N_t^{RF} = N_r^{RF} = 4$. From this figure we observe that MIB-SOMP algorithm exhibits same spectral efficiency as the SOMP algorithm, this is because the MIB-SOMP algorithm only changes the method of matrix inversion, but the hybrid precoding matrix is the same as that of the SOMP algorithm. The spectral efficiency of the RM-SVD algorithm is higher than that of the SOMP algorithm, which is closer to the optimal full-digital precoding algorithm. This algorithm better extracts the information in the optimization target matrix by performing singular value decomposition on the residual matrix. The OBMP algorithm has the lowest spectral efficiency, because the candidate matrix is defined in advance without CSI.

In Fig. 3(b), we compare the spectral efficiency of all hybrid precoding algorithms when the number of RF chains varies from 2 to 6. This simulation sets $N_s = 2$ and SNR $= -5$dB. As the number of RF chains increases, the spectral efficiency of the SOMP algorithm and improved hybrid precoding algorithm is closer to the optimal full-digital precoding algorithm. With the same SNR, the spectral efficiency of the RM-SVD algorithm is about 5%–10% higher than that of SOMP algorithm. And the spectral efficiency of the MIB-SOMP algorithm is the same as the spectrum efficiency of the SOMP algorithm. The OBMP algorithm has the lowest spectrum efficiency, which is about 80% of that of the SOMP algorithm.



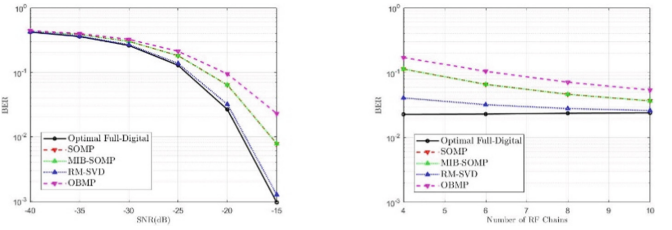
(a) Spectral efficiency versus SNR (b) Spectral efficiency versus the number of RF chains

Fig. 3. Spectral efficiency of hybrid precoding algorithms

4.3 Bit Error Rate

In Fig. 4(a), we compare the bit error rate (BER) of the algorithms. This simulation sets $N_s = 2$ and $N_t^{RF} = N_r^{RF} = 4$. Consistent with the spectral efficiency of algorithms, MIB-SOMP and SOMP have the same BER. The RM-SVD has a lower BER than the SOMP algorithm. The OBMP algorithm has the highest BER.

In Fig. 4(b), we compare the BER of the algorithms when the number of RF chains varies from 4 to 10. This simulation sets $N_s = 4$ and SNR $= -15$ dB. With the same SNR, we observe that the BER of the OBMP algorithm is increased by about 35%–40%, the BER of the RM-SVD algorithm is reduced by about 30%–40%. And the BER of the MIB-SOMP algorithm is the same as that of the SOMP algorithm.



(a) Bit error rate versus SNR (b) Bit error rate versus the number of RF chains

Fig. 4. Bit error rate of hybrid precoding algorithms

5 Conclusions

In this paper, we analyze the main issues of the SOMP algorithm, which are high complexity of matrix inversion and high dependence on the channel estimation. On this basis, we compare and analyze the performance and complexity of the three improved hybrid precoding algorithms. The RM-SVD algorithm is more complicated than the SOMP algorithm due to the SVD of the residual matrix, but the spectrum efficiency is approximately improved 5%–10%, and the bit error rate is approximately reduced

30%–40%. The OBMP algorithm uses the DFT codebook as a candidate matrix to completely avoid matrix inversion. The complexity is about 20% of the MIB-SOMP algorithm, the spectral efficiency is reduced by about 20%, and the bit error rate is increased by about 35%–40%. The MIB-SOMP algorithm uses the Schur-Banachiewicz blockwise inversion and replaces the matrix inversion with iterative inversion. The algorithm exhibits same performance as the SOMP algorithm via simulation results with reduced computational complexity.

References

1. Huang, H., Liu, K.P., Wen, R.: Joint channel estimation and beamforming for millimeter wave cellular system. In: IEEE Global Communications Conference, San Diego (2015)
2. Doan, C.H., Emami, S., Sobel, D.A., Niknejad, A.M.: Design considerations for 60 GHz CMOS radios. IEEE Commun. Mag. **42**, 132–140 (2004)
3. Vu, M., Paulraj, A.: MIMO wireless linear precoding. IEEE Signal Process. Mag. **24**, 86–105 (2007)
4. Marzetta, T.L.: Noncooperative cellular wireless with unlimited numbers of base station antennas. IEEE Trans. Wirel. Commun. **9**(11), 3590–3600 (2010)
5. Ayach, O.E., Rajagopal, S., Abu-Surra, S.: Spatially sparse precoding in millimeter wave MIMO systems. IEEE Trans. Wirel. Commun. **13**(3), 1499–1513 (2014)
6. Xiang, J.W., Yu, X.L., Jing, X.R.: Low complexity hybrid precoding method in mmWave massive MIMO system. Telecommun. Sci. **32**(09), 10–15 (2016)
7. Chen, C.H., Tsai, C.R., Liu, Y.H.: Compressive sensing assisted low-complexity beamspace hybrid precoding for millimeter-wave MIMO systems. IEEE Trans. Signal Process. **65**(6), 1412–1424 (2017)
8. Lee, Y.Y., Wang, C.H.: A hybrid RF/baseband precoding processor based on parallel-index-selection matrix-inversion-bypass simultaneous orthogonal matching pursuit for millimeter wave MIMO systems. IEEE Trans. Signal Process. **63**(2), 305–317 (2015)
9. Ayach, O.E., Heath, R.W., Abu-Surra, S.: Low complexity precoding for large millimeter wave MIMO systems. In: IEEE International Conference Communications (ICC), pp. 3724–3729 (2012)
10. Alkhateeb, A., El Ayach, O., Leus, G., Heath, R.W.: Heath: Hybrid precoding for millimeter wave cellular systems with partial channel knowledge. In: Information Theory and Applications Workshop (ITA), San Diego, CA, USA (2013)