Low Complexity Hybrid Precoding in Millimeter Wave Massive MIMO Systems

Tongtong Cheng[1*], Yigang He[2*], Yuting Wu[2], Shuguang Ning[2], Yongbo Sui[2] and Yuan Huang[2]
[1] China Academy of Information and Communications Technology
Haidian, Beijing 100191 China
[e-mail: chengtongtong@caict.ac.cn]
[2] Hefei University of Technology
Hefei, Anhui 230000 China
[e-mail: 18655136887@163.com]
[*]Corresponding author: Tongtong Cheng, Yigang He

Received July 19, 2021;revised March 8, 2022; accepted March 18, 2022; published April 30, 2022

Abstract

As a preprocessing operation of transmitter antennas, the hybrid precoding is restricted by the limited computing resources of the transmitter. Therefore, this paper proposes a novel hybrid precoding that guarantees the communication efficiency with low complexity and a fast computational speed. First, the analog and digital precoding matrix is derived from the maximum eigenvectors of the channel matrix in the sub-connected architecture to maximize the communication rate. Second, the extended power iteration (EPI) is utilized to obtain the maximum eigenvalues and their eigenvectors of the channel matrix, which reduces the computational complexity caused by the singular value decomposition (SVD). Third, the Aitken acceleration method is utilized to further improve the convergence rate of the EPI algorithm. Finally, the hybrid precoding based on the EPI method and the Aitken acceleration algorithm is evaluated in millimeter-wave (mmWave) massive multiple-input and multiple-output (MIMO) systems. The experimental results show that the proposed method can reduce the computational complexity with the high performance in mmWave massive MIMO systems. The method has the wide application prospect in future wireless communication systems.

Keywords: Power iteration, Aitken acceleration, extended power iteration, hybrid precoding, millimeter wave, massive MIMO systems.

This work was supported by the 2020 Industrial Internet Innovation and Development Project: Network Identifier Construction Project.

http://doi.org/10.3837/tiis.2022.04.013 **ISSN : 1976-7277**

1. INTRODUCTION

As the key technology of fifth generation mobile communication (5G), the massive multiple-input and multiple-output (MIMO) systems have hundreds of antennas to improve the spectral efficiency [1][2]. The millimeter wave (mmWave) ensures the feasibility of the massive MIMO systems in practical communication systems. Due to the limited wavelength of the millimeter waves, the Antenna in Package (AIP) technology can integrate hundreds or even thousands of system antennas into smaller chips or circuit boards [3]. Therefore, the combination of mmWave and massive MIMO systems is a promising technology.

In mmWave massive MIMO systems, the precoding technology is used to enhance the strength of the desired directional signal through the spatial processing to increase the system capacity [4]. The precoding technology adjusts the phase or amplitude of each element in the antenna array to obtain high transmission efficiency. However, the transmitter antennas of mmWave massive MIMO systems require many RF chains [5][6], which greatly increases the system complexity and the power consumption cost of the equipment [7][8]. Therefore, the hybrid precoding technology is the main solution in mmWave massive MIMO systems.

Based on the connection method, the hybrid precoding technology has three main categories. The first one is the fully-connected architectures-based hybrid precoding [9-13], and many scholars have made some researches about it. Such as, Ayach proposed sparse precoding, which transforms the achievable rate optimization problem into a sparse approximation problem and solves it by the matching pursuit algorithm, and its computation complexity is high [9]. Wonil et al. exploited codebooks to optimize the hybrid precoding by an iterative algorithm. The fully-connected architecture leads to a high computational complexity for iterative algorithms [10]. An alternative minimization algorithm based on the manifold optimization was introduced in [11]. The algorithm has the approximately optimal performances, but due to the manifold optimization, the digital and analog precoding matrixes are updated iteratively, so the computation complexity is still very high. All designed precoding for hybrid precoding employ a fully-connected architecture causing three additional limitations: 1) more phase shifters are required; 2) more energy is required to compensate for the energy consumption of the phase shifters; and 3) the computational complexity is higher, which leads to more energy consumption [9-14].

The second one is the sub-connected architecture-based hybrid precoding [15-16]. The each radio frequency (RF) chain is fixed to a certain portion of the antennas and can consume less power than the fully-connected structure. The successive interference cancelation (SIC) precoding has been proposed in sub-connected architectures [15]. The hybrid precoding optimization problem with non-convex constraints is transferred into a series of simple sub-antenna array optimization problems, each of which only considers one sub antenna array. However, the SIC precoding is a local-optimum for the sub-antenna array. The transmission efficiency is very limited. The hybrid precoding based on the alternating minimization algorithm has the similar drawback [16]. A hybrid precoding algorithm based on particle swarm optimization (PSO) has been introduced. The hybrid precoding uses the PSO algorithm to search the elements in the analog precoding matrix column by column. However, the method reduces the system performance to some extent [17]. Maximizing the spectral efficiency of the system is regarded as the objective function under the sub-connected architecture, and the optimization process of the analog precoding and the digital precoding are separated in two stages. However, the performance is still low [17-19]. Thus, the main concern of the sub-connected architecture is to improve the transmission performance.

The last one is the adaptive-connected structures-based hybrid precoding [20]. The

adaptive-connected structure exploits antenna-selected technology to connect subset RF chains, which can consume less power. Some studies have exploited the dynamic connection technology to select RF chain and antenna arrays [21-24]. [23-25] proposed the hybrid precoding based on a sub-connected architecture, which uses a dynamic connection algorithm to select sub-arrays. The adaptive-connected structure can provide a trade-off between the transmission speed and the power consumption by additional switched and antenna selection algorithms, which require more system resources. In addition, some previous studies on the hybrid precoding only investigated how to maximize the spectral efficiency. However, the computational complexity is the key in the hybrid precoding of the wireless communication systems, especially in 5G systems.

To address the problems of high complexity and limited performance, a sub-connected architecture with the low complexity hybrid precoding is proposed in this paper. To obtain well transmission performance and avoid the local optimization problem, the extended power iteration (EPI) method is utilized to calculate the maximum eigenvalues and their eigenvectors of the channel matrix. The eigenvectors of the EPI method are used to obtain the hybrid precoding matrix with low computation complexity. The contributions of this paper can be summarized as follows.

1) The EPI method is proposed to calculate the maximum eigenvalues and their eigenvectors. The power iteration (PI) is used to calculate the maximum eigenvalue and its eigenvectors of the channel matrix. Then the deflation method is used to obtain the maximum eigenvalues and their eigenvectors. Based on the EPI method, the calculation speed of the hybrid precoding is improved.

2) The Aitken method is used to accelerate the convergence of the EPI method. When the maximum eigenvalue is close to the second eigenvalue, the convergence is very slow. Therefore, the Aitken method is used to solve this problem. Additionally, the Aitken method can increase the convergence speed of the EPI method.

3) Experimental results obtained with the proposed precoding are given for mmWave massive MIMO systems. The results show that the hybrid precoding based on the EPI method and the Aitken method has low complexity and a faster calculation speed than other methods.

The remainder of this paper is organized as follows. Section 2 briefly describes the mathematical model of mmWave massive MIMO systems. In Section 3, the hybrid precoding based on the EPI method is introduced in detail. The experimental results are provided in Section 4. The conclusions are given in Section 5.

2. MATHEMATICAL MODEL OF HYBRID PRECODING FOR MMWAVE MASSIVE MIMO

2.1 System Model

A sub-connected point-to-point architecture mmWave MIMO systems including one base station (BS) with MN_{RF} antennas transmits N data streams to receivers with K antennas are considered in our paper, as shown in **Fig. 1**. [8]. The BS has N_{RF} RF chains and $N \le N_{RF} < K < MN_{RF}$. Therefore the hybrid precoding is divided into two parts: a $N_{RF} \times N_{RF}$ baseband precoder $\mathbf{F}_{BB}$ using N transmit chains and an $MN_{RF} \times N_{RF}$ RF precoder $\mathbf{F}_{RF}$ using analog circuits. When we transmit *N* data streams, the vectors $\mathbf{f}_{(N_{RF}-N)}$ and $f^{B}_{(N_{RF}-N)}$ are zero vectors. Thus, the received signal can be modelled as

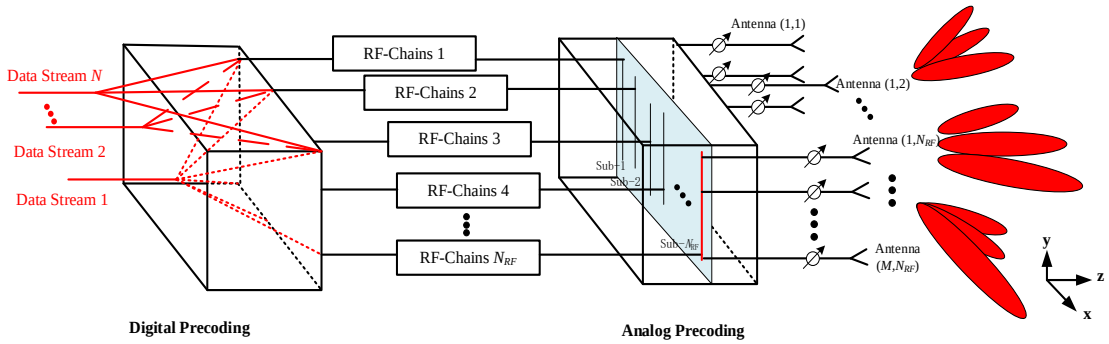


Fig. 1. Sub-connected architecture for mmWave massive MIMO systems.

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

$$\mathbf{F}_{RF} = \begin{bmatrix} \mathbf{f}_1 & 0 & 0 & 0 \\ \mathrm{M} & \mathrm{O} & \mathrm{O} & \mathrm{M} \\ 0 & 0 & \mathbf{f}_N & \mathrm{M} \\ 0 & 0 & \mathrm{L} & \mathbf{f}_{(N_{RF}-N)} \end{bmatrix}_{MN_{RF}\times N_{RF}} \tag{3}$$

where $\mathbf{H} \in \mathbb{C}^{K\times MN_{RF}}$ is the matrix of the complex channel gain; $\mathbf{s}$ is a $N\times 1$ symbol vector with $\mathrm{E}[\mathbf{s}\mathbf{s}^H] = \mathbf{I}_N / N$, $\mathbf{F} = \mathbf{F}_{RF}\mathbf{F}_{BB} = diag[\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_N, \mathbf{f}_{(N_{RF}-N)}] \cdot diag[f_1^B, f_2^B, ..., f_N^B, f_{(N_{RF}-N)}^B]$ is a $MN_{RF}\times N$ transmit hybrid precoding matrix in the sub-connected architecture; and $\mathbf{F}_{RF}$ is an analog precoding matrix, which controls the direction of the transmitter antennas. $\mathbf{F}_{BB}$ is the digital precoding, which controls the power of the transmitting antenna. $\mathbf{n}$ is the additive white Gaussian noise and $\gamma = P/\|\mathbf{F}\|_F^2$ is a scaling factor to satisfy the transmitting power constraint. Compared to the fully-connected architecture with MN_{RF}^2 phase shifters (PS), the sub-connection architecture requires MN_{RF} PSs.

In our system model, the extended Saleh Valenzuela cluster channel model is utilized to describe the mmWave channel characteristics [26-28]. In this channel model, the channel matrix **H** is assumed to be the sum of the contributions of L scattering clusters. Therefore, the channel matrix can be written as

$$\mathbf{H} = \sqrt{\frac{MN_{RF}K}{L}} \sum_{l=1}^{L} \alpha_l \Lambda_r(\phi_l^r, \theta_l^r) \Lambda_t(\phi_l^t, \theta_l^t) \mathbf{a}_r(\phi_l^r, \theta_l^r) \mathbf{a}_t(\phi_l^t, \theta_l^t) \tag{4}$$

where α_l denotes the complex gain of the lth path; $\Lambda_r(\phi_l^r, \theta_l^r)$ and $\Lambda_t(\phi_l^t, \theta_l^t)$ denote the antenna element gain for the arrival and departure angles, respectively [26]; and ϕ^t (ϕ^r) and θ^t (θ^r) denote the azimuth and elevation departure angles (arrival), respectively. $\mathbf{a}_r(\phi_l^r, \theta_l^r)$ and $\mathbf{a}_t(\phi_l^t, \theta_l^t)$ represent the receiving and transmitting antenna array response vectors, respectively.

The function of the antenna element gain $\Lambda(\phi, \theta)$ can be expressed as follows:

$$\Lambda(\phi, \theta) = \begin{cases} 1 & \forall \phi \in [\phi_{\min}, \phi_{\max}], \forall \theta \in [\theta_{\min}, \theta_{\max}] \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where $[\phi_{\min}, \phi_{\max}]$ denote the maximum and minimum azimuth angles of the receiving and transmitting antennas respectively. $[\theta_{\min}, \theta_{\max}]$ denote the maximum and minimum elevation angles, respectively.

For the array response vector (ARV), the two commonly-used antenna arrays are expressed as follows:

For the uniform linear array (ULA), the ARV with *M* elements is given by

$$\mathbf{a}_{ULA}(\phi) = \frac{1}{\sqrt{M}}[1, e^{j\frac{2\pi}{\lambda}d\sin(\phi)}, ..., e^{j(M-1)\frac{2\pi}{\lambda}d\sin(\phi)}] \tag{6}$$

In practical systems, the azimuth angle is quantized as an array antenna as follows:

$$\phi = \left\{ \phi_{\min} + \frac{\phi_{\max} - \phi_{\min}}{Q}, \phi_{\min} + 2\frac{\phi_{\max} - \phi_{\min}}{Q}, ..., \phi_{\max} - \frac{\phi_{\max} - \phi_{\min}}{Q} \right\} \tag{7}$$

where *Q* denotes the number of phase shifter quantization levels.

For a uniform planar array (UPA) with W_1 and W_2 elements on horizon and vertical, respectively, the ARV is expressed as follows:

$$\mathbf{a}_{UPA}(\phi,\theta) = \frac{1}{\sqrt{W_1 W_2}}\left[1, e^{j\frac{2\pi}{\lambda}d(w_1\sin(\phi)\sin(\theta)+w_2\cos(\theta))}, ..., e^{j\frac{2\pi}{\lambda}d((W_1-1)\sin(\phi)\sin(\theta)+(W_2-1)\cos(\theta))}\right] \tag{8}$$

where $0 \le w_1 \le W_1, 0 \le w_2 \le W_2$. (ϕ,θ) is quantized as follows:

$$\phi = \left\{ \phi_{\min} + \frac{\phi_{\max} - \phi_{\min}}{Q_1}, \phi_{\min} + 2\frac{\phi_{\max} - \phi_{\min}}{Q_1}, ..., \phi_{\max} - \frac{\phi_{\max} - \phi_{\min}}{Q_1} \right\} \tag{7}$$

$$\theta = \left\{ \theta_{\min} + \frac{\theta_{\max} - \theta_{\min}}{Q_2}, \phi_{\min} + 2\frac{\theta_{\max} - \theta_{\min}}{Q_2}, ..., \theta_{\max} - \frac{\theta_{\max} - \theta_{\min}}{Q_2} \right\} \tag{10}$$

where Q_1 and Q_2 denote the number of phase shifter quantization levels.

2.2 Hybrid Precoding Design

We consider the design of the hybrid precoding for point-to-point mmWave massive MIMO systems. For the sub-connected architecture, the achievable rate can be expressed as follows:

$$R = \log_2\left(\left|\mathbf{I} + \frac{\gamma}{N}\mathbf{H}\mathbf{F}_{RF}\mathbf{F}_{BB} \times \mathbf{F}_{BB}^H\mathbf{F}_{RF}^H\mathbf{H}^H\right|\right) \tag{11}.$$

To obtain the maximum achievable rate, $\mathbf{F}_{RF}, \mathbf{F}_{BB}$ are designed under the following constraints:

A) The total transmit power constraint should satisfy $\left\|\mathbf{F}_{RF}\mathbf{F}_{BB}\right\| \le N$

B) The PS requires that the analogy vectors have the same amplitude: $\mathbf{f}_i \in \mathbf{A}$.

where $\mathbf{A} = [\mathbf{a}(\phi_1,\theta_1), \mathbf{a}(\phi_2,\theta_2), ..., \mathbf{a}(\phi_i,\theta_i), \mathbf{a}(\phi_L,\theta_L)]$, $\mathbf{a}(\phi_i,\theta_i)$ is an $M \times 1$ ARV, and all ARVs are set as a similar matrix for every RF chain.

However, the non-convex constraints on $\mathbf{F}_{RF}\mathbf{F}_{BB}$ to maximize the achievable rates are complicated. Therefore, in some studies, the achievable rate has been maximized by reducing the "distance" between $\mathbf{F}_{RF}\mathbf{F}_{BB}$ and the optimal unconstrained precoding matrix. For hybrid architectures, many studies have proved that the design of the analog precoding matrix and the digital precoding matrix is developed from the optimal unconstrained precoding matrix by using sparse signal processing techniques. Ayach proved that the right-singular vectors of the channel are the optimal unconstrained vectors due to the anti-interference ability of SVD

precoding [9], which inspired us to make $\mathbf{F}_{RF}\mathbf{F}_{BB}$ approach the digital SVD precoding matrix in a sub-connected architecture. The optimal unconstrained precoding matrix can divide the channel into an independent channel to reduce the interference between the transmit antennas. In real mmWave communication ($N \le N_{RF} \le MN_{RF}$), when we transmit N data streams by N RF-chain, the hybrid precoding design can be decomposed as follows:

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H \tag{12}$$

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_1 & 0 \\ 0 & \mathbf{\Sigma}_2 \end{bmatrix}, \mathbf{V} = [\mathbf{V}_1, \mathbf{V}_2] \tag{13}$$

where $\mathbf{\Sigma}_1$ is $N \times N$ and $\mathbf{V}_1 = [\mathbf{v}_1, \mathbf{v}_2, ...\mathbf{v}_i, ...\mathbf{v}_N]$.

Due to the special sub-connected architecture and the independence of the vectors in $\mathbf{V}_1$, we can split the right-singular matrix $\mathbf{V}_1$ into N vectors. Then we exploit the N RF chain and N antenna sub-arrays to approach the independence vector $\mathbf{v}_i$ from $\mathbf{V}_1$.

When $f_i^B \mathbf{f}_i$ is close to $\mathbf{v}_i$,

$$\begin{aligned} \left\| \mathbf{v}_i - f_i^B \mathbf{f}_i \right\|_2^2 &= \left(\mathbf{v}_i - f_i^B \mathbf{f}_i \right)^H \left(\mathbf{v}_i - f_i^B \mathbf{f}_i \right) \\ &= \mathbf{v}_i^H \mathbf{v}_i + (f_i^B)^2 \mathbf{f}_i^H \mathbf{f}_i - 2 f_i^B \operatorname{Re}(\mathbf{v}_i^H \mathbf{f}_i) \\ &= 1 + (f_i^B)^2 - 2 f_i^B \operatorname{Re}(\mathbf{v}_i^H \mathbf{f}_i) \\ &= 1 + (f_i^B)^2 - 2 f_i^B \operatorname{Re}(\mathbf{v}_i^H \mathbf{f}_i) \\ &= \left(f_i^B - \operatorname{Re}(\mathbf{v}_i^H \mathbf{f}_i) \right)^2 + \left(1 - \operatorname{Re}^2(\mathbf{v}_i^H \mathbf{f}_i) \right) \end{aligned} \tag{14}$$

Because $\mathbf{V}_i$ is a unitary matrix and $\mathbf{f}_i$ has the same power allocation, we obtain $\mathbf{F}_{RF}, \mathbf{F}_{BB}$

$$\mathbf{f}_i = \frac{1}{\sqrt{M}} j\text{angle}(\mathbf{v}_1) \tag{15}$$

$$f_i^B = \frac{\|\mathbf{v}_i\|}{\sqrt{M}}. \tag{16}$$

In practical systems, the ARV is quantized. Therefore, the elements of $\mathbf{F}_{RF}$ are derived from $\mathbf{A}$; thus,

$$\mathbf{f}_i = \arg\max_{\mathbf{a}^{(m)} \in \mathbf{A}} \left\| \text{angle}(\mathbf{v}_i) \text{angle}(\mathbf{a}^{(m)}) \right\| \tag{17}$$

From (13), (16), and (17), the new hybrid precoding must calculate the N largest eigenvalues and their eigenvectors. It is necessary to design $\mathbf{F}_{RF}, \mathbf{F}_{BB}$. In addition, the rank of the mmWave channel is equal to that of L scattering clusters; if $L < N$, we can only transmit L data streams to ensure that the independent eigenvectors guarantee the communication performance. It should be noted that the complexity of SVD procedures is substantial when MN_{RF} is large, and so only the N largest eigenvalues and their eigenvectors are used. Therefore, a low computation complexity method is used to solve these constraints.

3. EXTENDED POWER ITERATION PRECODING

To avoid the SVD procedure, we must identify an approximation method to obtain the N largest eigenvalues and their eigenvectors of the channel matrix. This paper proposes a novel EPI method to solve this problem. Before the EPI method, the PI method is given as follows.

3.1 Power Iteration Method

Let $\mathbf{A} \in \mathbb{C}^{n \times n}$ be a matrix with eigenvalues [29][30]:

$$|\lambda_1| > |\lambda_2| \geq \ldots \geq |\lambda_n| \tag{18}$$

where λ_1 is a maximum eigenvalue, then we assume that $\mathbf{A}$ has a set of independent eigenvectors $\mathbf{X_1}, \mathbf{X}_2, \ldots, \mathbf{X}_n$ of unit length associated with the eigenvalue.

Assume an initial vector $\mathbf{p}_0 = \mathbf{p}$. The power method forms the sequence of vectors $\mathbf{p}_1 = \mathbf{Ap}$, $\mathbf{p}_2 = \mathbf{A}^2\mathbf{p}$,…, and the recursion can be expressed as follows:

$$\mathbf{p}_k = \mathbf{A}\mathbf{p}_{k-1} = \mathbf{A}^2\mathbf{p}_{k-2} = \mathbf{A}^3\mathbf{p}_{k-3} = \mathbf{A}^k\mathbf{p}_0 \tag{19}$$

Then, by expanding the initial vector along the eigenvectors $\mathbf{p}_0 = \sum_{j=1}^{n} \alpha_j \mathbf{x}_j$ and assuming that $\alpha_1 \neq 0$, we obtain $\mathbf{p}_k$ for $k = 1, 2, 3, \ldots$.

$$\begin{aligned} \mathbf{p}_k &= \mathbf{A}^k\mathbf{p}_0 = \sum_{j=1}^{n} \lambda_j^k \alpha_j \mathbf{x}_j \\ &= \lambda_1^k \left(\alpha_1 \mathbf{x}_1 + \sum_{j=2}^{n} \left(\frac{\lambda_j}{\lambda_1} \right)^k \alpha_j \mathbf{x}_j \right) \\ &= \lambda_1^k \alpha_1 \mathbf{x}_1 + \mathrm{O}\left(\left| \frac{\lambda_2}{\lambda_1} \right|^k \right) \end{aligned} \tag{20}$$

In practice, the l_2-norm of $\mathbf{p}_0$ is set as 1 to avoid overflow or underflow, and the initial recursion is modified as follows:

$$\begin{cases} \mathbf{p}_k = \mathbf{A}\mathbf{v}_{k-1} \\ m_k = \max(\mathbf{p}_k), k = 1, 2, 3\ldots \\ \mathbf{p}_k = \mathbf{p}_k / m_k \end{cases} \tag{21}$$

Due to $|\lambda_1| > |\lambda_2|$, when $k \to \infty$,

$$\begin{cases} \lambda_1 \approx m_k \\ \mathbf{x}_1 \approx \mathbf{p}_k \end{cases} \tag{22}$$

The PI method yields the maximum singular value λ_1 and the first right eigenvectors $\mathbf{x}_1$ by using (22). However, this PI method computes only the largest eigenvalue and its corresponding eigenvector. Moreover, the sparse precoding needs the N largest eigenvalues and their eigenvectors. Therefore, the EPI method is proposed to calculate the maximum N eigenvalues and their eigenvectors.

3.2 Hybrid Precoding Based on the EPI Method

Because PI only can calculate the maximum value, it is difficult to meet the hybrid precoding requirements. Therefore, we derive the EPI method from the PI method by using the deflation method. Based on the PI method, we delete the maximum eigenvalues and their eigenvectors from the matrix through the deflation method. Then, we obtain new sets of maximum eigenvalues and their eigenvectors from the new matrix. The EPI method for the channel matrix is presented as follows:

Step 1: Calculate the first maximum eigenvalue and eigenvector of $\mathbf{H}^{(i)}$ as follows:

$$\mathbf{z}^{(n\text{-}1)} = \mathbf{H}^{(i)}\mathbf{u}^{(n-1)} \tag{23}$$

$$\lambda^{(n\text{-}1)} = \arg\max\left|z_i^{(n\text{-}1)}\right| \tag{24}$$

$$\mathbf{u}^{(n)} = \frac{\mathbf{z}^{(n\text{-}1)}}{\lambda^{(n\text{-}1)}}, n < n' \tag{25}$$

Step 2: Deflate the channel matrix $\mathbf{H}^{(i)}$ by the deflation method:

$$\mathbf{H}^{(i+1)} = \mathbf{H}^{(i)} - \frac{\lambda_i^{(n)}}{\left\|\mathbf{u}_i^{(n)}\right\|_2^2}\left(\mathbf{u}_i^{(n)}\right)^T \mathbf{u}_i^{(n)}. \tag{26}$$

When a new matrix $\mathbf{H}^{(i+1)}$ is obtained, skip back to step 1. The eigenvalue and eigenvector of $\mathbf{H}^{(i+1)}$ are calculated using (24) and (26). The proofs of the deflation method in (26) are presented in the APPENDIX I.

3.3 EPI Method Optimization Based on the Aitken Method

Based on (20), although the EPI method is convergent, its convergence rate may be slow when $|\lambda_1| \approx |\lambda_2|$. Therefore, the Aitken method [31][32] is used to accelerate the convergence rate. We can describe the overall procedure of the Aitken iterations as follows:

$$(1)\ x_{k+1} = \varphi(x_k) \tag{27}$$

$$(2)\ x_{k+2} \approx \varphi(x_{k+1}) \tag{28}$$

$$(3)\ x_{k+1} \approx \frac{\left(x_{k+2}x_k - x_{k+1}^2\right)}{x_{k+2} - 2x_{k+1} + x_k} \tag{29}$$

The geometric interpretation of the Aitken acceleration method is shown in **Fig. 2**. Suppose that x_0 is an approximate root of equation $y = g(x)$, where $x_1^{(1)} = g(x_0)$, $x_1^{(2)} = g(x_1^{(1)})$. The point pair can give $P_0(x_0, x_1^{(1)})$ and $P_1(x_1^{(1)}, x_1^{(2)})$, and the cross point of the chain P_0P_1 and function $y = x$ is x_1, which comes from the Aitken method. The proofs of the Aitken method are presented in APPRODIX II.

By the Aitken method, the (24) can be transformed to

$$\lambda^n = \begin{cases} m^{(n)} = \arg\max\left|z_i^{(n)}\right|, & \text{for } 1 \le n \le 2 \\ \dfrac{m^{(n)}m^{(n-2)} - (m^{(n-1)})^2}{m^{(n)} - 2m^{(n-1)} + m^{(n-2)}}, & \text{elsewise} \end{cases} \tag{30}$$

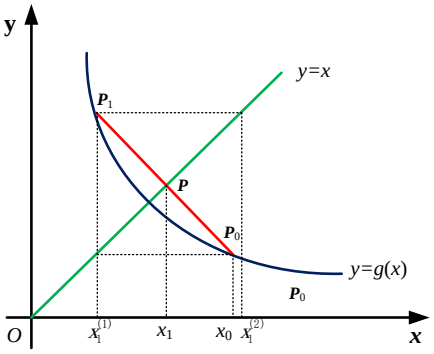


Fig. 2. Geometric interpretation of the Aitken method.

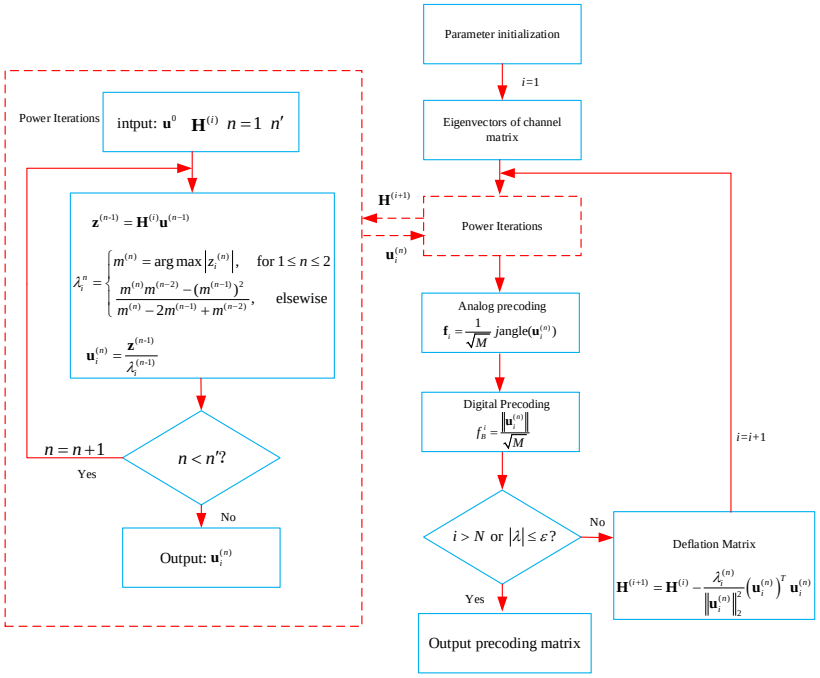


Fig. 3. A block diagram of EPI hybrid precoding

3.4 Procedure of Hybrid Precoding Based on EPI Method

Through N iterations, we can obtain the largest eigenvalue and the corresponding eigenvector of the channel matrix $\mathbf{H}$, $\mathbf{F}_{BB}$ and $\mathbf{F}_{RF}$ are accepted from the EPI method. Those specific steps in the EPI method are as follows in **Algorithm 1** and **Fig. 3**.

Algorithm 1. Expansion power iteration method

Require: $\mathbf{H}^{(1)}=\mathbf{H}$, $\mathbf{u}^0$, n', N, M $\mathbf{A}$

1 : Do :

2 : **For** $1\le n\le n'$

3 : $\mathbf{z}^{(n)}=\mathbf{H}^{(i)}\mathbf{u}^{(n-1)}$, $m^{(n)}=\arg\max\left|z^{(n)}\right|$

4 : **If** $1\le n\le 2$ **:** $\lambda_i^{(n)}=m^{(n)}$.

5 : **Else:** $\lambda_i^{(n)}=\dfrac{m^{(n)}m^{(n-2)}-(m^{(n-1)})^2}{m^{(n)}-2m^{(n-1)}+m^{(n-2)}}$, $\mathbf{u}_i^{(n)}=\dfrac{\mathbf{z}^{(n)}}{\lambda_i^{(n)}}$.

6 : **End for**

7 : $f_i^B=\left\|\mathbf{u}_i^{(n)}\right\|/\sqrt{M}$

8 : $\mathbf{f}_i=\underset{\mathbf{A}^{(m)}\in\mathbf{A}}{\arg\max}\left\|\text{angle}(\mathbf{u}_i^{(n)})\text{angle}(\mathbf{A}^{(m)})\right\|$

9: $\mathbf{H}^{(i+1)}=\mathbf{H}^{(i)}-\dfrac{\lambda_i^{(n)}}{\left\|\mathbf{u}_i^{(n)}\right\|_2}\left(\mathbf{u}_i^{(n)}\right)^T\mathbf{u}_i^{(n)}$

10: **Until** $i>N$ or $\left|\lambda_i^{(n)}\right|\le\varepsilon$?

11: **Return** $\mathbf{F}_{BB}$, $\mathbf{F}_{RF}$

3.5 Complexity Analysis

In this section, we compare the computational complexity of the SVD, the EPI, the dynamic sub-array (DS)[23], and the sparse precoding. The number of floating point operations (flops) needed to implement these complex operations varies greatly according to the used hardware and the complex number representation [33]. Our EPI precoding needs to compute the optimal victors $\mathbf{u}^{(n)}$ through the EPI method according to [31]. The matrix multiplication operation of $\mathbf{z}^{(n)} = \mathbf{H}^{(i)}\mathbf{u}^{(n-1)}$ and the Aitken method is $O\left(MN_{RF}K\right)$. In addition, the division operation of the EPI method is $O\left(2n'-2\right)$. Therefore, the total operation of the EPI method is $O(NN_{RF}MKn')+O\left(2n'N\right)$. the SVD and the DS precoding have similar computation complexity, i.e., $O((N_{RF}M)^3)$. The computation complexity of the sparse precoding method is $O(N_{RF}^4M + N_{RF}^2L^2 + N_{RF}^2M^2L)$. All hybrid precoding complexities are shown in **Table 1**.

In mmWave systems, the number of iterations *n* and the number of RF chains N are less than NM, M, and L. and the number of BS antennas NM is far larger than the other parameters. The critical parameter in the sparse precoding complexity is NM. Therefore, the SVD of the sparse precoding is the main contributor to the computational complexity. As shown in **Fig. 4**, the complexity of the EPI method is lower than that of the sparse precoding.

Table 1. Complexity of five hybrid precoding in mmWave massive MIMO systems

Precoding Algorithm	Complexity
Sparse[9]	$O(N_{RF}^4M + N_{RF}^2L^2 + N_{RF}^2M^2L)$
EPI	$O(NN_{RF}MKn')$
SIC[15]	$O(M^2(N_{RF}n' + K))$
DS[23]	$O(N_{RF}^3M^3)$
SVD	$O(N_{RF}^3M^3)$

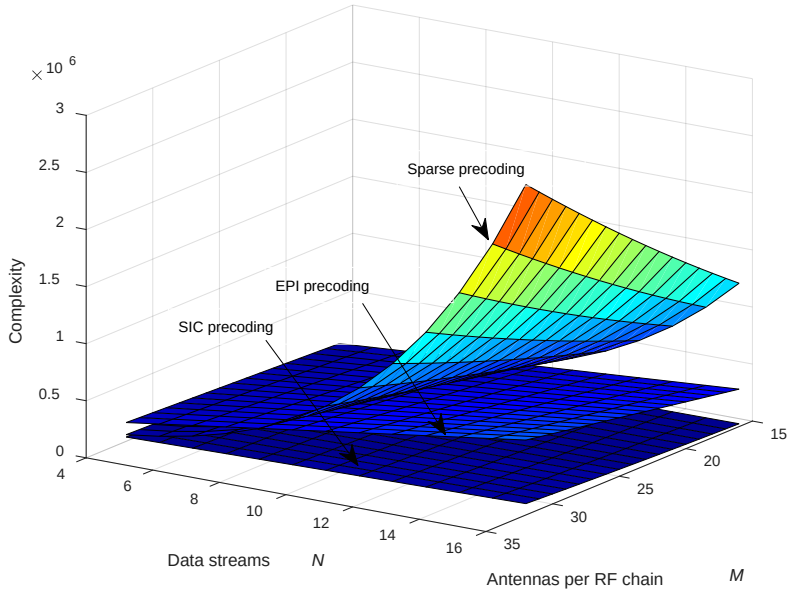


Fig. 4. The complexity comparisons of sparse, EPI, and SIC precoding for a (L = 32, n=5) mmWave MIMO systems

Fig. 4 shows the complexity of antennas per RF chain and RF chains for mmWave MIMO systems, considering mmWave massive MIMO systems with L=32, n' =5, K=16, and N_{RF}=16. Based on **Fig. 4**, the SIC precoding has the lowest complexity, but the lowest achievable rate cannot satisfy the communication requirement. The EPI precoding has a lower complexity than the sparse precoding. The precoding complexity has a light relationship with M; the main complexity is determined by the number of data streams N. In other words, when we increase the number of antennas per RF chain, the computation complexity has limited influence, but it can improve the achievable rate.

4. SIMULATION AND ANALYSIS

To demonstrate the performance of the hybrid precoding, we consider the classical mmWave massive MIMO systems. We use the software defined ratio (SDR), the channel emulation (CE), and the upconverter for our simulation experiment. The CE is used to generate real communication channel information with 8 XILINX KINTEX7 FPGAs; the modulation and the EPI precoding are deployed in the SDR systems. The data stream is upconverted into CE, and all experimental data is analyzed in MATLAB 2014a on a personal computer with 3.3 GHz quad-core processor with 16 GB of RAM. The simulation parameters are shown in **Table 2**. The ULA is used for simulation with the angle of departure (AoD) randomly distributes in $[-\pi/6, \pi/6]$ and the direction of departure (DoD) randomly distributes in $[-\pi, \pi]$. Therefore, we obtaine the ARV from (5).

Table 2. The experiment simulation parameter

Simulation parameters	Number
Effective channel paths L	32
Carrier frequency	28GHz
PI iteration	5
Antennas per RF chain	[16:4:32]
Number of data streams N	[4:16]
Receiver antenna K	16
AoDs of AVR	$[-\pi/6, \pi/6]$
DoD of AVR	$[-\pi, \pi]$
Quantization levels	2^5

4.1 Analysis of Achievable Rate with Antennas per RF Chain M.

In this section, the experimental results of the antennas per RF chain are analyzed and compared with the results of five different precoding techniques. **Fig. 5** and **Fig. 6** show the achievable rates of the SVD precoding, the sparse precoding, the DS precoding, the EPI precoding, and the SIC precoding with different M and signal-to-noise ratio (SNR) values.

As shown in **Fig. 5**, the number of RF chains N is set to 8. We obtaine different achievable rates with various M. The fully digital SVD precoding has the best performance comparing to the other precoding methods under all SNRs and BS antennas. The sparse precoding has the next highest performance and the DS precoding shows a performance similar to that of the sparse precoding. The SIC precoding has low performance due to the local-optimum of the sub-connected architecture. The proposed precoding has a similar performance to the sparse precoding. Due to the sparsity of the massive MIMO channel matrix, the maximum

eigenvalues and their eigenvectors represent the main components of the channel matrix. Thus, the EPI precoding has a better performance by using the top eigenvalues and their eigenvectors.

Fig. 6 shows the achievable rate of five precodings with MN=128, N=8, and SNR=0 dB. Similar to **Fig. 5**. the performance in descending order is the SVD precoding, the sparse precoding, the DS precoding, the EPI precoding, and the SIC precoding. However, the EPI precoding is close to the performance of the SVD precoding when the SNR value is large, but the computation complexity of the EPI precoding is lower than that of the sparse precoding.

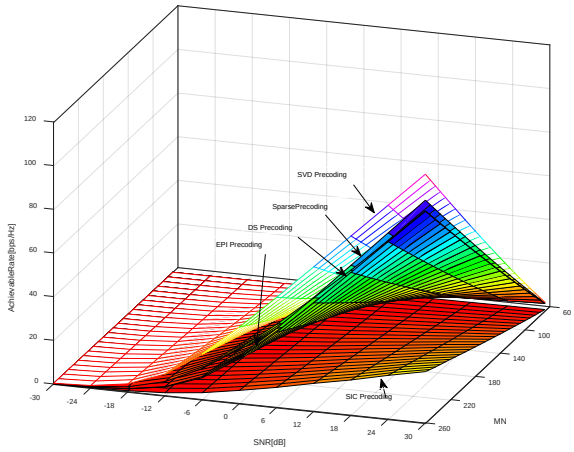


Fig. 5. Achievable rate of five precoding methods with the number of transmitter per RF chain antennas and SNR(N=8).

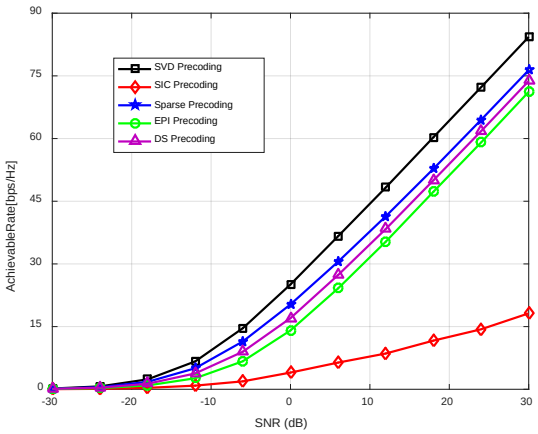


Fig. 6. Achievable rate comparison for 128×16 mmWave MIMO systems (N=8).

4.2 Analysis of Achievable Rate with Data Streams N.

In this section, the experimental results of the data streams N are analyzed with the result five different precoding. **Fig. 7** and **Fig. 8** show the achievable rate of five precoding when M=16. The performances are like that shown in **Fig. 5**, the EPI precoding is close to the performance of the sparse precoding and the DS precoding when SNR>0 dB. The achievable rate of the sparse precoding and the EPI precoding reach 102 bps/Hz when the SNR is 30 dB. Under a low SNR environment, those five precoding has small performance gaps due to high noise power, which limits the total achievable rate. Especially in **Fig. 8**, as the SNR increasing, all

precoding had a better achievable rate. The “gap” between the sparse precoding and the EPI precoding decreases as the SNR increases. However, the EPI precoding has a lower cost and computation complexity than the sparse precoding.

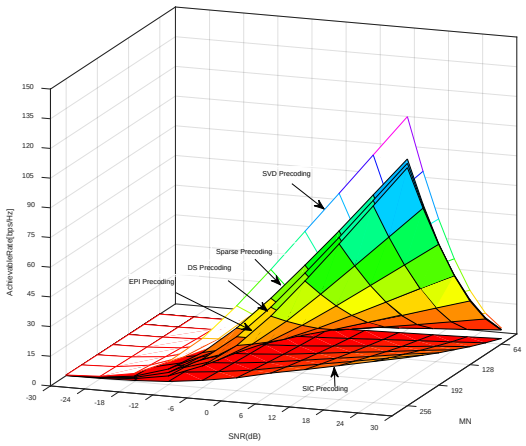


Fig. 7. Achievable rate of five precoding methods with different data stream N and SNR(M=16).

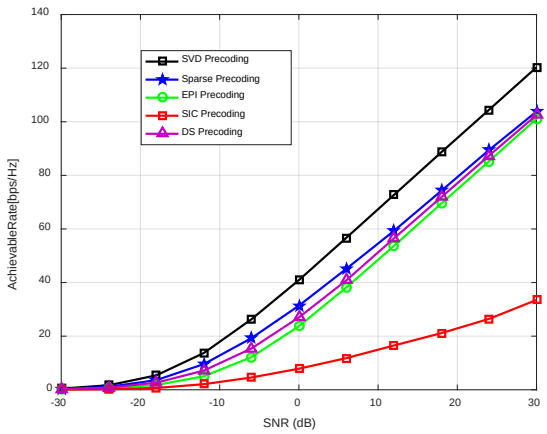


Fig. 8. Achievable rate comparison for 256×16 mmWave MIMO systems (M=16).

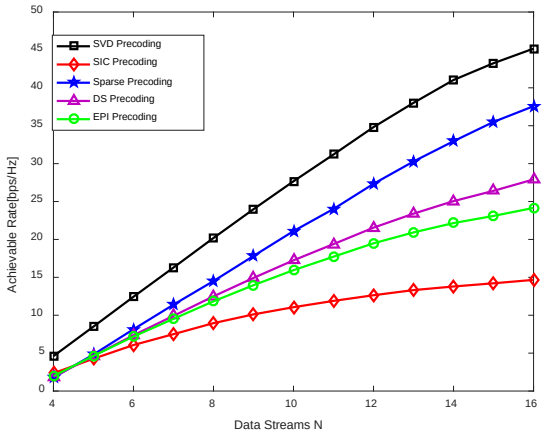


Fig. 9. Achievable rate comparison against the number of the data streams (SNR=0dB, L=32).

Table 3. The performance comparison of five precodings

precoding	Architecture	RF chain	PSs	Complexity	Achievable Rate
SVD	Digital	256	0	2×10^7	64.56bps/Hz
DS	dynamic subarray	16	256	2×10^7	51.06 bps/Hz
Sparse	fully-connected	16	4096	1.0×10^6	52.15bps/Hz
EPI	sub-connected	16	256	3.2×10^5	48.54bps/Hz
SIC	sub-connected	16	256	2.5×10^4	15.12bps/Hz

Fig. 9 shows the achievable rate variation against the number of different data streams N, when $M=16$ and SNR = 0 dB. Based on the **Fig. 9**, the performance of the EPI precoding is worse than that of the sparse precoding, the DS precoding, and the digital SVD precoding. The fully-connected architecture hybrid precoding has the better performance than the DS and the sub-connected architecture. However, the sub-connected architecture has a lower complexity and cost.

To compare the cost and complexity of the different precoding, the performance comparisons of those five precoding are given in **Table 3**. As shown in **Table 3**, the four hybrid precoding architectures including the full-digital (SVD precoding), the full-connected (sparse precoding), the dynamic subarray (DS precoding) and the sub-connected (EPI and SIC precoding) architectures, the RF chains, the PSs, the complexity, and the achievable rate are compared for the mmWave massive MIMO systems with M=16, N_{RF}=16, L=32, n=5, SNR=10 dB and N=16. As shown in **Table 3**, the SVD precoding required 256 RF chains, and the RF chain and PS are costly for communication systems; thus the cost of the SVD precoding is immense. Although the fully-connected architecture requires the similar RF chains, it needs more PSs than the dynamic subarray and sub-connected architecture. the DS precoding has a better performance than the EPI precoding, but its computational complexity is enormous. Although the EPI precoding and SIC precoding are based on a sub-connected architecture and the complexity of the EPI precoding is greater than that of the SIC precoding, the achievable rate of the EPI precoding (45.84bps/Hz) is greater than that of the SIC precoding (15.12bps/Hz) with sub-connected architecture. Overall, the EPI precoding provides a trade-off between transmission rate, the computation complexity and the cost. Thus, in a real communication environment, the hybrid precoding based on the EPI method is suitable for mmWave communication systems.

4.3 Analysis of Energy Efficiency

The energy efficiency of the hybrid precoding is expressed as the ratio of the achievable rate and energy consumption to express the trade-off between these two factors [25]:

$$E \triangleq \frac{R}{P_{tot}} \tag{31}$$

where P_{tot} is the total power consumption of the transmitter. The total power consumption for different hybrid precoding architectures is defined in **Table 4**.

Table 4. Transmit power consumption for different architecture in mmWave massive MIMO systems

Architecture	Total power consumptions
Full-connected	$P + 2P_{BB} + MNP_{RF} + MN^2 P_{PS}$
Dynamic- sub-connected	$P + 2P_{BB} + MNP_{RF} + MN^2 P_{PS} + MNP_{sw}$
Sub-connected	$P + 2P_{BB} + MNP_{RF} + MNP_{PS}$

P is the transmit power, P_{BB} is the digital power, P_{RF} is the RF power, P_{PS} is the PS power, and P_{SW} is the switch network power. In our simulation, P is 10w, P_{BB} is 200mW, P_{RF} is 300mW, P_{PS} is 50mW, and P_{SW} is 5mW. The energy efficiency of different precoding is given in **Fig. 10**. Due to the dynamic subarray architecture, the DS precoding has the best energy efficiency but its complexity is immense compared to that of other precoding. Additionally, the sub-connected architecture has a better energy efficiency than the sparse precoding with a fully-connected architecture. The EPI precoding has a better energy performance than the SIC precoding and the sparse precoding. Therefore, the proposed low complexity precoding is more energy efficient for mmWave massive MIMO systems.

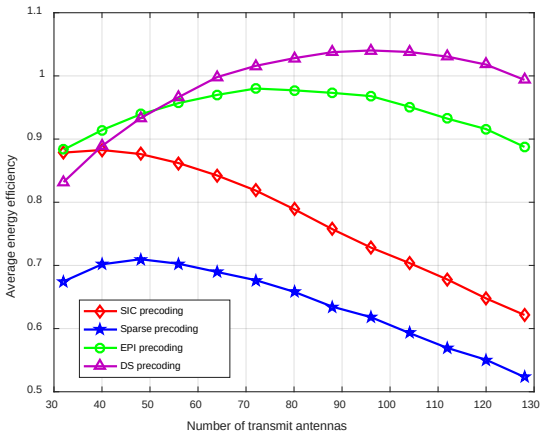


Fig. 10. Average energy efficiency for different architecture in 128×16 mmWave MIMO system.

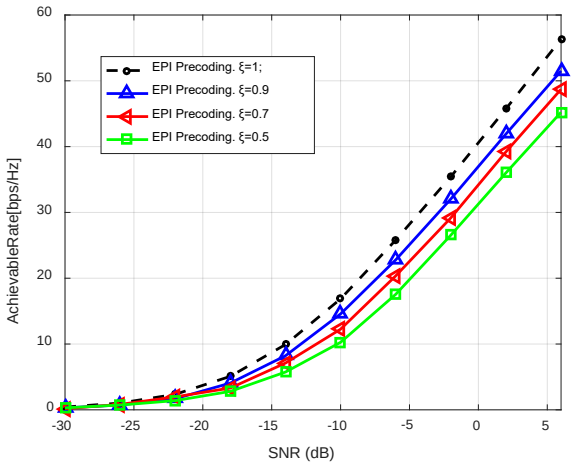


Fig. 11. Achievable rate comparison of different ξ for an $MN \times K = 256\times16$ ($M = 16$) mmWave MIMO system.

4.4 Analysis of Achievable Rate with Imperfect CSI.

The effect of the imperfect CSI is evaluated with the EPI hybrid precoding. The imperfect CSI $\hat{\mathbf{H}}$ can be modelled as [1]

$$\hat{\mathbf{H}} = \xi \mathbf{H} + \sqrt{1-\xi^2}\mathbf{E} \tag{32}$$

where $\mathbf{H}$ is the perfect channel matrix, $\xi \in [0,1]$ is the CSI accuracy, and $\mathbf{E}$ is the error matrix with the distribution **i.i.d**. CN (0, 1). **Fig. 11** shows the sum achievable rate comparison for an $MN \times K = 256 \times 16$ (N =8) mmWave MIMO system, for which different ξ scenarios are simulated. The EPI hybrid precoding is robust to different ξ. As shown in **Fig. 11**, the achievable rate of the EPI precoding approaches the perfect CSI scenario (ξ = 1) when ξ = 0.9. Even when the CSI accuracy is quite poor (i.e., ξ = 0.5), the EPI precoding achieves more than 80% sparse precoding performance in the perfect CSI scenario. Therefore, the EPI precoding has the better performance in the imperfect CSI scenario.

4.5 Analysis of Achievable Rate with Scatting Cluster L.

The rank of the mmWave channel is equal to the number of scattering clusters L; if L is small, the number of data streams is N, and the performance of the hybrid precoding decreases, because the independent eigenvectors of the channel matrix cannot support transmitting N data streams. To explain the relationship between the spectral effectiveness and the number of scattering clusters L, the simulation results are shown in **Fig. 12**. The number of data streams is set to 16. When the scattering clusters are small at 16, the achievable rate increases as L increases because when $L<N$, the number of eigenvectors cannot support the transmission of N data streams, which provides L independence eigenvectors to design hybrid precoding. When $L>N$, the performance of the hybrid precoding is not limited by the eigenvectors of the channel matrix. The number of data streams N influences the achievable rate. As shown in **Fig. 9**, when L=32, the achievable rate increases as N increases.

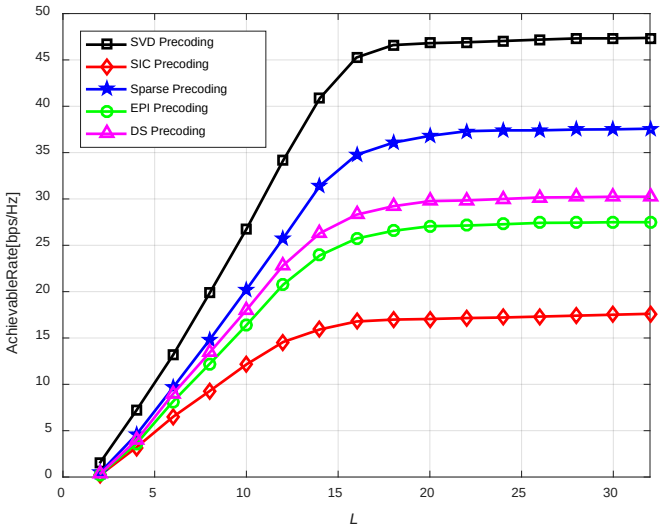


Fig. 12. Achievable rate comparison of different scattering clusters mmWave MIMO systems (M =16, N=16, and SNR=0dB).

5. Conclusion

In this paper, we propose a novel hybrid precoding for mmWave massive MIMO systems. First, we utilize the hybrid precoding to approach the maximum eigenvectors of the channel matrix, with the aim of obtaining the maximum achievable rate in the sub-connected architecture. Then we introduce the EPI method to obtain the maximum eigenvalues and their eigenvectors to avoid the SVD procedure. Then the Aitken method is used to accelerate the convergence of the EPI method. Based on the EPI and Aitken methods, we can apply hybrid precoding in mmWave massive MIMO systems with low computation complexity. Finally, the simulation results show that the EPI algorithm has a near-optimal performance approaching sparse precoding and digital SVD precoding under low levels of computational complexity.

In addition, some further investigations will focus on two directions. 1) The impacts of the SNR on our EPI method will be studied; and we will conduct deeper research on filter technology before the hybrid precoding. 2) Quantitative analysis will be given more attention between lower complexity and faster computational speed in our upcoming work.

Appendix

APPENDIX I (EPI method)

Let $\mathbf{B} = \mathbf{A} - \lambda_1 \mathbf{I}$, Then $(\lambda_i - \lambda_1, \mathbf{p}_i)$ are eigenpairs of $\mathbf{B}$.

Proof:

$$\begin{aligned} \mathbf{B}\mathbf{p}_1 &= \mathbf{A}\mathbf{p}_1 - \lambda_1 \mathbf{p}_1 \\ &= \lambda_1 \mathbf{p}_1 - \lambda_1 \mathbf{p}_1 \\ &= 0\mathbf{p}_1 \end{aligned} \tag{33}$$

and

$$\begin{aligned} \mathbf{B}\mathbf{p}_i &= \mathbf{A}\mathbf{p}_i - \lambda_1 \mathbf{p}_i \\ &= \lambda_i \mathbf{p}_i - \lambda_1 \mathbf{p}_i \\ &= (\lambda_i - \lambda_1)\mathbf{p}_i \end{aligned} \tag{34}$$

Assuming $\mathbf{A}$ is a Hermitian matrix, the matrices $\mathbf{A}$ and $\mathbf{A}^H$ have the same set of eigenvalues. Let $(\lambda_i, \mathbf{w}_i)$ and $(\lambda_j, \mathbf{p}_j)$ denote the eigenpairs of $\mathbf{A}^H$ and $\mathbf{A}$. If $\lambda_i \neq \lambda_j$, $\mathbf{w}_i^T \mathbf{p}_j = 0$.

Proof:

$$\begin{aligned} 0 &= \mathbf{w}_i^T (\mathbf{A}\mathbf{p}_j) - \mathbf{p}_i^T (\mathbf{A}^T \mathbf{w}_j) \\ &= \lambda_j \mathbf{w}_i^T \mathbf{p}_j - \lambda_i \mathbf{p}_j^T \mathbf{w}_i \\ &= (\lambda_j - \lambda_i)\mathbf{w}_i^T \mathbf{p}_j \end{aligned} \tag{35}$$

If $\lambda_i \neq \lambda_j$, $\mathbf{w}_i^T \mathbf{p}_j = 0$.

Let $\mathbf{B} = \mathbf{A} - \lambda_i \mathbf{p}_i \mathbf{x}^T$ where $\mathbf{p}_i^T \mathbf{x} = 1$. Then the eigenvalues of $\mathbf{B}$ are 0, and λ_j for $j \neq i$.

Proof:

$$\begin{aligned}\mathbf{B}\mathbf{p}_i &= \mathbf{A}\mathbf{p}_i - \lambda_i \mathbf{p}_i \mathbf{x}^T \mathbf{p}_i \\ &= \lambda_i \mathbf{p}_i - \lambda_i \mathbf{p}_i \\ &= 0\end{aligned} \tag{36}$$

and

$$\begin{aligned}\mathbf{B}^T \mathbf{w}_j &= (\mathbf{A} - \lambda_i \mathbf{p}_i \mathbf{x}^T)^T \mathbf{w}_j \\ &= \mathbf{A}^T \mathbf{w}_j - \lambda_i \mathbf{x}\mathbf{p}_i^T \mathbf{w}_j \\ &= \mathbf{A}^T \mathbf{w}_j \\ &= \lambda_j \mathbf{w}_j\end{aligned} . \tag{37}$$

Based on (35), we can use the PI method to obtain eigenvalues by deflation method as follows:

$$\mathbf{x}_i = \frac{\mathbf{u}_i}{\|\mathbf{u}_i\|_2^2}, \mathbf{u}_1 = \mathbf{p}_1 \tag{38}$$

$$\mathbf{B}_{(i+1)} = \mathbf{B}_{(i)} - \lambda_i \mathbf{u}_i \mathbf{x}_i^T \tag{39}$$

APPENDIX II (Aitken method)

According to the traditional differential mean value theorem:

$$x_{k+1} = \varphi(x_k) \tag{40}$$

$$x - x_{k+1} = \varphi(x) - \varphi(x_k) = \varphi'(\xi)(x - x_k) . \tag{41}$$

Assume. There is a constant $|q| < 1$ such that

$$x - x_{k+1} \approx q(x - x_k) , \tag{42}$$

then

$$x_{k+1} \approx \varphi(x_{k+1}) , x - x_{k+1} = q(x - x_{k+1}) . \tag{43}$$

$$\frac{x - x_{k+1}}{x - x_{k+1}} = \frac{(x - x_k)}{(x - x_{k+1})} \tag{44}$$

$$x \approx x_{k+1} - \frac{(x_{k+1} - x_{k+1})^2}{x_{k+1} - 2x_{k+1} + x_k} \tag{45}$$

$$x_{k+1} \approx \frac{(x_{k+1} x_k - x_{k+1}^2)}{x_{k+1} - 2x_{k+1} + x_k} \tag{46}$$

References

[1] F. Rusek et al., “Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays,” *IEEE Signal Proce. Mag.*, vol. 30, no. 1, pp. 40-60, Jan. 2013. Article(CrossRef Link)

[2] R. Zi, X. Ge, J. Thompson, C.-X. Wang, H. Wang, and T. Han, “Energy effciency optimization of 5G radio frequency chain systems,” *IEEE J. Sel.Areas Commun.*, vol. 34, no. 4, pp. 758-771, Apr. 2016. Article (CrossRef Link)

[3] J. Park, D. Choi and W. Hong, "Millimeter-Wave Phased-Array Antenna-in-Package (AiP) Using Stamped Metal Process for Enhanced Heat Dissipation," *IEEE Antennas and Wireless Propagation Letters*, vol. 18, no. 11, pp. 2355-2359, Nov. 2019. Article (CrossRef Link)

[4] A. Alkhateeb, J. Mo, N. González-Prelcic, and R. Heath, "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014. Article (CrossRef Link)
[5] W. Roh et al., "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb. 2014. Article (CrossRef Link)
[6] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp.101–107, June. 2011. Article (CrossRef Link)
[7] W. Zhang, X. Xia, Y. Fu and X. Bao, "Hybrid and full-digital beamforming in mmWave Massive MIMO systems: A comparison considering low-resolution ADCs," *China Commun.*, vol. 16, no. 6, pp. 91-102, June 2019. Article (CrossRef Link)
[8] Q. Ding, Y. Deng, X. Gao and M. Liu, "Hybrid precoding for mmWave massive MIMO systems with different antenna arrays," *China Commun.*, vol. 16, no. 10, pp. 45-55, Oct. 2019. Article (CrossRef Link)
[9] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wire. Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014. Article (CrossRef Link)
[10] C. Chen, "An Iterative Hybrid Transceiver Design Algorithm for Millimeter Wave MIMO Systems," *IEEE Wire. Commun. Let.*, vol. 4, no. 3, pp. 285-288, June 2015. Article (CrossRef Link)
[11] X. Yu, J. Shen, J. Zhang and K. B. Letaief, "Alternating Minimization Algorithms for Hybrid Precoding in Millimeter Wave MIMO Systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 485-500, April 2016. Article (CrossRef Link)
[12] J. Jin, Y. R. Zheng, W. Chen and C. Xiao, "Hybrid Precoding for Millimeter Wave MIMO Systems: A Matrix Factorization Approach," *IEEE Trans. Wire. Commun.*, vol. 17, no. 5, pp. 3327-3339, May 2018. Article (CrossRef Link)
[13] L. Liang, W. Xu and X. Dong, "Low-Complexity Hybrid Precoding in Massive Multiuser MIMO Systems," *IEEE Wire. Commun. Lett.*, vol. 3, no. 6, pp. 653-656, Dec. 2014. Article (CrossRef Link)
[14] W.U. Khan, X. Li, A. Ihsan, et al., "NOMA-enabled Optimization Framework for Next-generation Small-cell IoV Networks under Imperfect SIC Decoding," *IEEE Trans. Intel. Trans. Sys.*, 2021. Artcle (CrossRef Link)
[15] X. Gao, L. Dai, S. Han, C. I and R. W. Heath, "Energy-Efficient Hybrid Analog and Digital Precoding for MmWave MIMO Systems With Large Antenna Arrays," *IEEE J. Sel.Areas Commun.*, vol. 34, no. 4, pp. 998-1009, April 2016. Article (CrossRef Link)
[16] J. Du, W. Xu, H. Shen, X. Dong and C. Zhao, "Hybrid Precoding Architecture for Massive Multiuser MIMO With Dissipation: Sub-Connected or Fully Connected Structures?," *IEEE Trans. Wire. Commun.*, vol. 17, no. 8, pp. 5465-5479, Aug. 2018. Article (CrossRef Link)
[17] O. Alluhaibi, Q. Z. Ahmed, C. Pan and H. Zhu, "Capacity Maximization for Hybrid Digital-to-Analog Beamforming mm-Wave Systems," in *Proc. of 2016 IEEE Global Communications Conference (GLOBECOM)*, Washington, DC, pp. 1-6, 2016. Article (CrossRef Link)
[18] W.U. Khan, F. Jameel, X. Li, et al., "Joint Spectrum and Energy Optimization of NOMA-enabled Small-Cell Networks with QoS Guarantee," *IEEE Trans. Vehic. Techn.*, vol. 70, no. 8, pp. 8337-8342, 2021. Article (CrossRef Link)
[19] W. Ye, Z. Weixia, "Hybrid precoding design for millimetre wave systems with the partially-connected structure," *IET Commun.*, vol.14, no.4, pp.561-567, 2020. Article (CrossRef Link)
[20] X. Xue, Y. Wang, L. Yang, et al., "Energy-Efficient Hybrid Precoding for Massive MIMO mmWave Systems with A Fully-Adaptive-Connected Structure," *IEEE Transactions on Communications*, vol. 68, no. 6, pp. 3521-3535, June 2020. Article (CrossRef Link)
[21] V. V. Ratnam, A. F. Molisch, O. Y. Bursalioglu, and H. C. Papadopoulos, "Hybrid beamforming with selection for multiuser massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 15, pp. 4105–4120, Aug. 2018. Article (CrossRef Link)

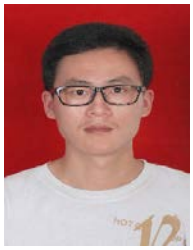
[22] Y. Gao, H. Vinck, and T. Kaiser, “Massive MIMO antenna selection: Switching architectures, capacity bounds, and optimal antenna selection algorithms,” *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1346–1360, Mar. 2018. Article (CrossRef Link)
[23] S. Park, A. Alkhateeb and R. W. Heath, "Dynamic Subarrays for Hybrid Precoding in Wideband mmWave MIMO Systems," *IEEE Trans. Wirel. Commun.*, vol. 16, no. 5, pp. 2907-2920, May 2017. Article (CrossRef Link)
[24] L. Yan, C. Han and J. Yuan, "A Dynamic Array-of-Subarrays Architecture and Hybrid Precoding Algorithms for Terahertz Wireless Communications," *IEEE J. Sel.Areas Commun.*, vol. 38, no. 9, pp. 2041-2056, Sept. 2020. Article (CrossRef Link)
[25] H. Li, M. Li, Q. Liu and A. L. Swindlehurst, "Dynamic Hybrid Beamforming With Low-Resolution PSs for Wideband mmWave MIMO-OFDM Systems," *IEEE J. Sel.Areas Commun.*, vol. 38, no. 9, pp. 2168-2181, Sept. 2020. Article (CrossRef Link)
[26] Z. Pi and F. Khan, “An introduction to millimeter-wave mobile broadband systems,” *IEEE Commun. Mag*., vol. 49, no. 6, pp. 101–107, Jun. 2011. Article (CrossRef Link)
[27] H. S.F., I. C.L., X. Z.K., and R. C., “Large-scale antenna systems with hybrid precoding analog and digital beamforming for millimeter wave 5G,” *IEEE Commun. Mag*., vol. 53, no. 1, pp. 186–194, Jan. 2015. Article (CrossRef Link)
[28] C.A. Balanis, *Antenna Theory: Analysis and Design*, Hoboken, NJ, USA: Wiley, 2012
[29] G. H. Golub and C. F. Van Loan, *Matrix Computations*, MD, USA: JHU Press, 2012.
[30] Å. Björck, *Numerical methods in matrix computations*, Berlin, German: Springer, 2015.
[31] Chen J., Gan M., Zhu Q., et al., “Robust Standard Gradient Descent Algorithm for ARX Models Using Aitken Acceleration Technique,” *IEEE Trans. Cybern.*, (early access), 2021. Article (CrossRef Link)
[32] T. X.Q., W. L.J., and F. W., *Numerical Calculation*, Beijing, China: Sci. Press, 2015.
[33] S. Boyd, L. Vandenberghe, “Numerical linear algebra background,”. Accessed 12 Feb 2016. [Online]. Available: num-lin-alg2.pdf (stanford.edu)

Tongtong Cheng received a bachelor degree in electrical engineering from Hefei University of Technology, Hefei, Anhui in 2016 and the Ph.D. degree in electrical engineering from Hefei University of Technology, Hefei, Anhui in 2021. He joined in China Academy of Information and Communications Technology in 2021. His research interest is wireless communication and Industrial Internet.

Yigang He received the M.Sc. degree in electrical engineering from Hunan University, Changsha, China, in 1992 and the Ph.D. degree in electrical engineering from Xi'an Jiaotong University, Xi'an,China, in 1996. In 1990, he joined the College of Electrical and Information Engineering, Hunan University and was promoted to Associate Professor, Professor in 1996, 1999, respectively. From 2006 to 2011, he worked as the Director of the Institute of Testing Technology for Circuits and Systems, Hunan University. He was a Senior Visiting Scholar with the University of Hertfordshire, Hatfield, U.K., in 2002. In 2011, he joined the Hefei University of Technology, China, and currently works as the Head of School of Electrical Engineering and Automation, Hefei University of Technology.

Yuting Wu received the B.S. degree in Hefei University of Technology, Anhui, China, in 2015. He is now purchasing the Ph.D. degree in electrical engineering from Hefei University of Technology, Hefei, China. His current interests include wireless communication technology, and smart grid technology.

Shuguang Ning received the M.Sc. degree in electrical theory and advanced technology from Hefei University of Technology, Hefei, China, in 2019, and is currently pursuing the Ph.D. degree in electrical engineering from Hefei University of Technology, Hefei, China. His current interests include the research of wireless communication, deep learning, intelligent perception and its application in power internet of things.



Yongbo Sui received a B.E. degree in electrical engineering from Xiangtan University, Xiangtan, Hunan, in 2014. He is currently working toward a PhD at Hefei University of Technology, Hefei, Anhui. His research interests include intelligent algorithms and wireless communication channels.

Yuan Huang received the B.S. degree in Hefei University of Technology, Anhui, China, in 2016. He is now pursuing the Ph.D. degree in electrical engineering from Hefei University of Technology, Hefei, China. His current interests include the research of digital signal processing, wireless communication, and compressed sensing