Hybrid Precoding/Combining Design in mmWave Amplify-and-Forward MIMO Relay Networks

Lisi Jiang, Xiaoyi (Leo) Liu, Hamid Jafarkhani

***Abstract*—In this paper, we consider the amplify-and-forward relay networks in mmWave systems and propose a hybrid precoder/combiner design approach. The phase-only RF precoding/combining matrices are first designed to support multi-stream transmission, where we compensate the phase for the eigenmodes of the channel. Then, the baseband precoders/combiners are performed to achieve the maximum mutual information. Based on the data processing inequality for the mutual information, we first jointly design the baseband source and relay nodes to maximize the mutual information before the destination baseband receiver. The proposed low-complexity iterative algorithm for the source and relay nodes is based on the equivalence between mutual information maximization and the weighted MMSE. After we obtain the optimal precoder and combiner for the source and relay nodes, we implement the MMSE-SIC filter at the baseband receiver to keep the mutual information unchanged, thus obtaining the optimal mutual information for the whole relay system. Simulation results show that our algorithm achieves better performance with lower complexity compared with other algorithms in the literature.**

I. Introduction

Communications over millimeter wave (mmWave) has received significant attention recently because of the high data rates providing by the large bandwidths at the mmWave carrier frequencies. Also, using large antenna arrays in mmWave communication systems is possible because the small wavelength allows integrating many antennas in a small area. Despite its advantages, the mmWave carrier frequencies suffer from relatively severe propagation losses. Meanwhile, the sparsity of the mmWave scattering environment usually results in rank-deficient channels [1].

To overcome the large path losses, large antenna arrays can be placed at both transmitters and receivers to guarantee sufficient received signal power [2]. The large antenna arrays lead to a large number of RF chains, which greatly increase the implementation cost and complexity. To reduce the number of RF chains, hybrid analog/digital precoding has been proposed, which connects analog phase shifters with a reduced number of RF chains. The main advantage of the hybrid precoding is that it can trade off between the low-complexity limited-performance analog phase shifters and the high-complexity good-performance digital precoding [3].

Despite the help of large antenna arrays, the severe propagation losses still limit mmWave communications to take place within short ranges. Fortunately, the coverage can be greatly extended with the help of relay nodes [4]. Therefore, investigating the performance of hybrid precoding/combining in the relay scenario is important. For conventional relay scenario, network beamforming in amplify-and-forward (AF) relay networks was studied in [5], and different relay selection methods have been proposed and analyzed in [6].

For a mmWave relay scenario, large antenna arrays are usually implemented to mitigate the severe path loss. In addition, a hybrid precoding method is adopted. There are two typical hybrid precoding structures: (i) fully-connected structure (where each RF chain is connected to all antennas) [7], and (ii) sub-connected structure (where each RF chain is connected to a subset of antennas) [8]. For fully-connected mmWave networks with AF relay nodes, the authors in [9] designed hybrid precoding matrices using the orthogonal matching pursuit (OMP) algorithm. However, the performance of the OMP algorithm used in [9] depends on the orthogonality of the pre-determined candidates for the analog precoders. For sub-connected structures, [10] proposes a minimum mean squared error (MMSE)-based relay hybrid precoding design. To make the problem tractable, [10] reformulates the original problem as three subproblems and proposes an iterative successive approximation (ISA) algorithm. The algorithm in [10] can also be extended to the fully-connected structure. Compared with the OMP algorithm, the ISA algorithm in [10] greatly improves the performance, however, the complexity of the ISA algorithm is high and it only optimizes the relay node.

In this paper, we study the hybrid precoding for fully-connected mmWave AF relay networks in the domain of massive MIMO. We first design the phase-only RF precoding/combining matrices for multi-stream transmissions. We decompose the channel into parallel sub-channels through singular value decomposition (SVD) and compensate the phase of each sub-channel, i.e., each eigenmode of the channel. When the RF precoding and combining are performed, the digital baseband precoders/combiners are performed on the equivalent baseband channel to achieve the maximal mutual information. The problem of finding the optimal baseband precoders/combiners for the optimal mutual information is non-convex and intractable to solve using low complexity methods. Based on the data processing inequality for the mutual information [11], we first jointly design the baseband source and relay nodes to maximize the mutual information before the destination baseband receiver. We propose a low-

The authors are with the Center for Pervasive Communications and Computing, UC, Irvine (email: {lisi.jiang, xiaoyil3, hamidj}@uci.edu). This work was supported in part by the NSF Award ECCS-1642536.

complexity iterative algorithm to design the precoder and combiner for the source and relay nodes, which is based on the equivalence between mutual information maximization and the weighted MMSE [12]. After we obtain the optimal precoder and combiner for the source and relay nodes, we implement the MMSE successive interference cancellation (MMSE-SIC) filter [13] at the baseband receiver to keep the mutual information unchanged, thus obtaining the optimal mutual information for the whole relay system. Simulation results show that our algorithm outperforms the OMP in [9]. Moreover, our algorithm achieves better performance with lower complexity compared to the ISA algorithm in [10].

II. System Model

In this section, we present the signal and channel model for a single user mmWave MIMO relay system with large antenna arrays and limited RF chains.

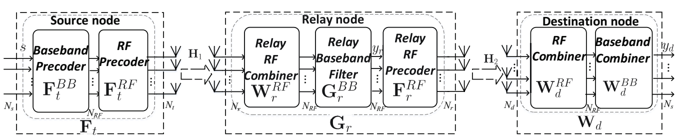

Fig. 1: System model

A. System Model

Consider a single-user mmWave MIMO relay system using hybrid precoding as illustrated in Fig. 1. The system consists of a source node with N_t transmission antennas, a relay node with N_r antennas for both transmitting and receiving signals, and a destination node with N_d antennas. Assuming N_s data streams are transmitted, the BS is equipped with N_{RF} RF chains such that $N_s \leq N_{RF} \leq N_t$. Using the N_{RF} transmit chains, an $N_{RF} \times N_s$ baseband precoder $\mathbf{F}_t^{BB}$ is applied. The RF precoder is an $N_t \times N_{RF}$ matrix $\mathbf{F}_t^{RF}$. Half duplex relaying is adopted. During the first time slot, the BS transmits the N_s data streams to the relay through a MIMO channel $\mathbf{H}_1 \in \mathbb{C}^{N_t \times N_r}$. The relay receives the signal with an RF combiner $\mathbf{W}_r^{RF} \in \mathbb{C}^{N_r \times N_{RF}}$ and a baseband filter $\mathbf{G}_r^{BB} \in \mathbb{C}^{N_{RF} \times N_{RF}}$. During the second time slot, the relay transmits the data using one RF percoder $\mathbf{F}_r^{RF} \in \mathbb{C}^{N_{RF} \times N_r}$ through a MIMO channel $\mathbf{H}_2 \in \mathbb{C}^{N_r \times N_d}$ and the destination receives the data with one RF combiner $\mathbf{W}_d^{RF} \in \mathbb{C}^{N_d \times N_{RF}}$ and one baseband combiner $\mathbf{W}_d^{BB} \in \mathbb{C}^{N_{RF} \times N_s}$.

We assume the transmited signal is $\boldsymbol{s} = [s^1, s^2, ..., s^{N_s}]^T$ with $\mathbb{E}[\boldsymbol{s}\boldsymbol{s}^H] = \mathbf{I}_{N_s} \in \mathbb{C}^{N_s \times N_s}$. During the first time slot, the received signal after the baseband filter at the relay can be expressed as

$$\boldsymbol{y}_r = \mathbf{G}_r^{BB}\mathbf{W}_r^{RF}\mathbf{H}_1\mathbf{F}_t^{RF}\mathbf{F}_t^{BB}\boldsymbol{s} + \mathbf{G}_r^{BB}\mathbf{W}_r^{RF}\boldsymbol{n}_1, \tag{1}$$

where $\boldsymbol{n}_1 \in \mathbb{C}^{N_r \times 1}$ is a zero-mean complex Gaussian noise vector at the relay node with covariance matrix $\mathbb{E}[\boldsymbol{n}_1\boldsymbol{n}_1^H] = \sigma_1^2\mathbf{I}_{N_r} \in \mathbb{C}^{N_r \times N_r}$. The power constraint at the source node is

$$\left\|\mathbf{F}_t^{RF}\mathbf{F}_t^{BB}\right\|_F^2 \leq E_t \tag{2}$$

During the second time slot, the received signal after the combiners at the destination can be expressed as

$$\begin{aligned}\boldsymbol{y}_d = &\mathbf{W}_d^{BB}\mathbf{W}_d^{RF}\mathbf{H}_2\mathbf{F}_r^{RF}\mathbf{G}_r^{BB}\mathbf{W}_r^{RF}\mathbf{H}_1\mathbf{F}_t^{RF}\mathbf{F}_t^{BB}\boldsymbol{s} \\ &+ \mathbf{W}_d^{BB}\mathbf{W}_d^{RF}\mathbf{H}_2\mathbf{F}_r^{RF}\mathbf{G}_r^{BB}\mathbf{W}_r^{RF}\boldsymbol{n}_1 + \mathbf{W}_d^{BB}\mathbf{W}_d^{RF}\boldsymbol{n}_2,\end{aligned} \tag{3}$$

where $\boldsymbol{n}_2 \in \mathbb{C}^{N_d \times 1}$ is a zero-mean complex Gaussian noise vector at the destination node with covariance matrix $\mathbb{E}[\boldsymbol{n}_2\boldsymbol{n}_2^H] = \sigma_2^2\mathbf{I}_{N_d} \in \mathbb{C}^{N_d \times N_d}$.

To simplify the expression, we define $\mathbf{F}_t = \mathbf{F}_t^{RF}\mathbf{F}_t^{BB} \in \mathbb{C}^{N_t \times N_s}$ as the hybrid precoding matrix at the transmitter, $\mathbf{G}_r = \mathbf{W}_r^{RF}\mathbf{G}_r^{BB}\mathbf{F}_r^{BB} \in \mathbb{C}^{N_r \times N_r}$ as the hybrid filter at the relay node, and $\mathbf{W}_d = \mathbf{W}_d^{BB}\mathbf{W}_d^{RF} \in \mathbb{C}^{N_s \times N_d}$ as the hybrid combiner at the destination node. Eq. (4) can be expressed as

$$\boldsymbol{y}_d = \mathbf{W}_d\mathbf{H}_2\mathbf{G}_r\mathbf{H}_1\mathbf{F}_t\boldsymbol{s} + \mathbf{W}_d\mathbf{H}_2\mathbf{G}_r\boldsymbol{n}_1 + \mathbf{W}_d\boldsymbol{n}_2, \tag{4}$$

The relay's power constraint is

$$\|\mathbf{G}_r\mathbf{H}_1\mathbf{F}_t\boldsymbol{s} + \mathbf{G}_r\boldsymbol{n}_1\|_F^2 \leq E_r \tag{5}$$

Based on this hybrid precoding/combining system model, we can derive the achieved data rate for the system

$$R = \frac{1}{2}\log_2\det(\mathbf{I}_{N_s} + \mathbf{W}_d\mathbf{H}_2\mathbf{G}_r\mathbf{H}_1\mathbf{F}_t\mathbf{R}_n^{-1}\mathbf{F}_t^H\mathbf{H}_1^H\mathbf{G}_r^H\mathbf{H}_2^H\mathbf{W}_d^H), \tag{6}$$

where $\mathbf{R}_n = \sigma_1^2\mathbf{W}_d\mathbf{H}_2\mathbf{G}_r\mathbf{G}_r^H\mathbf{H}_2^H\mathbf{W}_d^H + \sigma_2^2\mathbf{W}_d\mathbf{W}_d^H$ is the covariance matrix of the colored Gaussian noise at the output of the baseband combiner.

Generally, we want to jointly optimize the RF and baseband precoders/combiners to achieve optimal data rate. However, finding global optima for this problem (maxmizing R while constant-amplitude contraints imposed to the RF analog precoder/combiners) is non-convex and intractable. Separated RF and baseband processing designs, as [14] did, are investigated to obtain satisfying performance. Therefore, we will separate the RF and baseband domain designs in this paper.

B. Channel Model

MmWave channels are expected to have limited scattering characteristic [15], which means the assumptions of a rich scattering environment become invalid. This is called sparsity in the literature and leads to the unreliability of traditional channel models, such as the Rayleigh fading channel model. To characterize the limited scattering feature, we adopt the clustered mmWave channel model in [7] with L scatters. Each scatter is assumed to contribute a single propagation path. Then, the channel is given by

$$\mathbf{H} = \sqrt{\frac{N_tN_r}{L}}\sum_{l=1}^{L}\alpha_l\boldsymbol{a}_r(\varphi_l^r, \theta_l^r)\boldsymbol{a}_t^H(\varphi_l^t, \theta_l^t), \tag{7}$$

where L is the number of propagation paths, α_l is the complex gain of the l^{th} path, $\varphi_l^r(\theta_l^r)$ and $\varphi_l^t(\theta_l^t)$ are its random azimuth (elevation) angles of arrival and departure and $\boldsymbol{a}_r(\varphi_l^r, \theta_l^r)$ ($\boldsymbol{a}_t(\varphi_l^t, \theta_l^t)$) is the receiving (transmitting) antenna array response vectors. While the algorithms and results in the paper can be applied to arbitrary antenna arrays, we use uniform linear arrays (ULAs) in the simulations for simplicity. The array response vectors take the following form

$$\boldsymbol{a}^{ULA}(\varphi) = \frac{1}{\sqrt{N}}[1, e^{jkdsin(\varphi)}, ..., e^{j(N-1)kdsin(\varphi)}]^T, \tag{8}$$

where $k = \frac{2\pi}{\lambda}$ and d is the spacing between antenna elements. The angle φ is assumed to have a uniform distribution over $[0, 2\pi]$.

The channel can be further expressed as $\mathbf{H} = \mathbf{A}_r diag(\boldsymbol{\alpha})\mathbf{A}_t^H$, where $\mathbf{A}_r = [\boldsymbol{a}_r(\varphi_1^r, \theta_1^r), ..., \boldsymbol{a}_r(\varphi_L^r, \theta_L^r)]$ and $\mathbf{A}_t = [\boldsymbol{a}_t(\varphi_1^t, \theta_1^t), ..., \boldsymbol{a}_t(\varphi_L^t, \theta_L^t)]$ are matrices containing the receiving and transmitting array response vectors, and $\boldsymbol{\alpha} = [\alpha_1, ..., \alpha_L]^T$.

III. Hybrid Precoder/Combiner Design

As discussed in Section II, we use a hybrid design to reduce the number of RF chains. We first design the RF precoder/combiner. Then, based on the designed RF precoder/combiner, we design a low-complexity iterative algorithm for the baseband precoder/combiner to maximize the mutual information.

A. RF Precoder/Combiner Design

Our goal for RF precoder/combiner is to make the channels decomposed into N_{RF} parallel sub-channels to support the multi-stream transmission. The main challenge is the constant-magnitude constraints on RF precoders and combiners. Without the constant-magnitude constraints, the optimal precoder/combiner should be the right/left singular matrix of the channel, which transmit the signals along the eigenmodes of the channel. Considering the constant-magnitude constraints, we cannot directly use the singular matrix to rotate the signals, but we can use the projection on each eigenmode as a criterion to choose RF precoder and combiner. For the i^{th} eigenmode, the best precoder should be the one that has the largest projection on that eigenmode, i.e., the one that casts the most energy along that eigenmode direction.

Using $\mathbf{H}_1$ as an example, we first perform the singular decomposition for the channel matrix.

$$\mathbf{H}_1 = \mathbf{U}_1\Sigma_1\mathbf{V}_1^H = \sum_i^L \sigma_i \boldsymbol{u}_i\boldsymbol{v}_i^H, \tag{9}$$

where $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ are the i^{th} vectors in matrix $\mathbf{U}_1$ and $\mathbf{V}_1$, which correspond to σ_i. L is the rank of the channel, which equals to the number of propagation paths. Eq. (9) indicates that channel $\mathbf{H}_1$ has L eigenmodes. We denote the i^{th} eigenmode as $\boldsymbol{u}_i\boldsymbol{v}_i^H$, and it has a gain of σ_i.

For our RF precoding/combining, we want to maximize the projection of the i^{th} data stream onto the i^{th} eigenmode, i.e., $|\boldsymbol{w}_i^H\boldsymbol{u}_i\boldsymbol{v}_i^H\boldsymbol{f}_i|$, where $\boldsymbol{w}_i$ and $\boldsymbol{f}_i$ are the i^{th} vector of precoder $\mathbf{F}_t^{RF}$ and combiner $\mathbf{W}_r^{RF}$. To approach the maximal projection, we have the following proposition.

Proposition 1. *The optimal phase-only vectors $\boldsymbol{f}_i$ and $\boldsymbol{w}_i$, which maximize the projection for the i^{th} data stream onto the i^{th} eigenmode of the channel, will satisfy the following conditions:*

$$phase(\boldsymbol{f}_i[m]) = phase(\boldsymbol{v}_i[m]),\ \forall m = 1, 2, ..., N_t \tag{10}$$

$$phase(\boldsymbol{w}_i[n]) = phase(\boldsymbol{u}_i[n]),\ \forall n = 1, 2, ..., N_t \tag{11}$$

where $[k]$ represents the k^{th} element of a vector.

Proof. First, we express the vectors in polar coordinates. Due to the magnitude-constant constraints, vectors $\boldsymbol{f}_i$ and $\boldsymbol{w}_i$ are expressed as $\boldsymbol{f}_i = \frac{1}{\sqrt{N_t}}[e^{j\theta_1^i}, e^{j\theta_2^i}, ..., e^{j\theta_{N_t}^i}]^T$ and $\boldsymbol{w}_i = \frac{1}{\sqrt{N_r}}[e^{j\varphi_1^i}, e^{j\varphi_2^i}, ..., e^{j\varphi_{N_r}^i}]^T$. Since there are no constant-magnitude constraints for $\boldsymbol{v}_i$ and $\boldsymbol{u}_i$, each element in the vector has its own magnitude. The polar forms of $\boldsymbol{v}_i$ and $\boldsymbol{u}_i$ are $\boldsymbol{v}_i = [r_1^i e^{j\alpha_1^i}, r_2^i e^{j\alpha_2^i}, ..., r_{N_t}^i e^{j\alpha_{N_t}^i}]^T$ and $\boldsymbol{u}_i = [\rho_1^i e^{j\beta_1^i}, \rho_2^i e^{j\beta_2^i}, ..., \rho_{N_r}^i e^{j\beta_{N_r}^i}]^T$. Then, the projection can be calculated as

$$\left|\boldsymbol{w}_i^H\boldsymbol{u}_i\boldsymbol{v}_i^H\boldsymbol{f}_i\right| = \left|\frac{1}{\sqrt{N_t}}\sum_{n=1}^{N_r}\rho_n^i e^{j(\varphi_n^i-\beta_n^i)}\right|\left|\frac{1}{\sqrt{N_t}}\sum_{m=1}^{N_t} r_m^i e^{j(\alpha_m^i-\theta_m^i)}\right| \tag{12}$$

According to Cauchy-Schwartz inequality, we have

$$\left|\frac{1}{\sqrt{N_t}}\sum_{m=1}^{N_t} r_m^i e^{j(\alpha_m^i-\theta_m^i)}\right|^2 \leq \frac{1}{N_t}\sum_{m=1}^{N_t}\left|r_m^i\right|^2\sum_{m=1}^{N_t}\left|e^{j(\alpha_m^i-\theta_m^i)}\right|^2 = \frac{1}{N_t}\sum_{m=1}^{N_t}\left|r_m^i\right|^2. \tag{13}$$

Equality can be achieved in (13) if and only if $\theta_m^i = \alpha_m^i$, $\forall m = 1, 2, ..., N_t$. This means the maximal $|\boldsymbol{v}_i^H\boldsymbol{f}_i|$ is achieved when $\theta_m^i = \alpha_m^i$, $\forall m = 1, 2, ..., N_t$. Similarly, the maximal $|\boldsymbol{w}_i^H\boldsymbol{u}_i|$ is achieved when $\varphi_n^i = \beta_n^i$, $\forall n = 1, 2, ..., N_r$. Therefore, we have the conclusion that the optimal phase-only vectors $\boldsymbol{f}_i$ and $\boldsymbol{w}_i$, which maximize the $|\boldsymbol{w}_i^H\boldsymbol{u}_i\boldsymbol{v}_i^H\boldsymbol{f}_i|$, will satisfy the conditions in (10) and (11). □

Our RF precoders and combiners are actually compensating the phase of each sub-channel. Note that when the number of antennas is large enough, the response vectors $\boldsymbol{a}_r(\varphi_l^r, \theta_l^r)$s and $\boldsymbol{a}_t(\varphi_l^t, \theta_l^t)$s become orthogonal to each other. $\mathbf{A}_t$ and $\mathbf{A}_r$ become the left and right singular matrices of the channel and they directly become our RF precoder and combiner. In this case, we can perfectly decompose the channel into independent parallel sub-channels. The equivalent channel after RF processing is diagonal, which makes it easier for baseband processing.

B. Baseband system

In this section, we focus on designing the baseband precoding/combining matrices. First we define the equivalent baseband channels for $\mathbf{H}_1$ and $\mathbf{H}_2$ as

$$\tilde{\mathbf{H}}_1 = \mathbf{W}_r^{RF}\mathbf{H}_1\mathbf{F}_t^{RF}, \tag{14}$$

$$\tilde{\mathbf{H}}_2 = \mathbf{W}_d^{RF}\mathbf{H}_2\mathbf{F}_r^{RF}. \tag{15}$$

Based on the equivalent channels, we simplify our system model as shown in Fig. 2.

Using the equivalent channels (14) and (15), we rewrite the received signals at the destination node as

$$\tilde{\boldsymbol{y}}_d = \tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB}\boldsymbol{s} + \tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\boldsymbol{n}}_1 + \tilde{\boldsymbol{n}}_2 \tag{16}$$

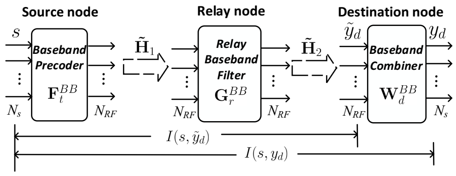

Fig. 2: Baseband System model

$$\boldsymbol{y}_d = \mathbf{W}_d^{BB}\tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB}\boldsymbol{s} + \mathbf{W}_d^{BB}\tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\boldsymbol{n}}_1 + \mathbf{W}_d^{BB}\tilde{\boldsymbol{n}}_2, \qquad (17)$$

where $\tilde{\boldsymbol{n}}_1 = \mathbf{W}_r^{RF}\boldsymbol{n}_1$ and $\tilde{\boldsymbol{n}}_2 = \mathbf{W}_d^{RF}\boldsymbol{n}_2$.

Our ultimate goal for the baseband design is to maximize the mutual information $I(\boldsymbol{s},\boldsymbol{y}_d)$. However, directly optimizing $I(\boldsymbol{s},\boldsymbol{y}_d)$ is intractable. According to the data processing inequality [11], i.e., $I(\boldsymbol{s},\boldsymbol{y}_d) \leq I(\boldsymbol{s},\tilde{\boldsymbol{y}}_d)$, we first design $\mathbf{F}_t^{BB}$ and $\mathbf{G}_r^{BB}$ to maximize the mutual information $I(\boldsymbol{s},\tilde{\boldsymbol{y}}_d)$. After we get the maximum $I(\boldsymbol{s},\tilde{\boldsymbol{y}}_d)$, we implement the MMSE-SIC for $\mathbf{W}_d^{BB}$, which according to [13] is information lossless. In this way, we make $I(\boldsymbol{s},\boldsymbol{y}_d) = I(\boldsymbol{s},\tilde{\boldsymbol{y}}_d)$. Since $I(\boldsymbol{s},\tilde{\boldsymbol{y}}_d)$ is maximized, $I(\boldsymbol{s},\boldsymbol{y}_d)$ is also maximized because of the data processing inequality and the independence of $I(\boldsymbol{s};\tilde{\boldsymbol{y}}_d)$ from $\mathbf{W}_d^{BB}$.

C. $\mathbf{F}_t^{BB}$ and $\mathbf{G}_r^{BB}$ design

In this section, we jointly design $\mathbf{F}_t^{BB}$ and $\mathbf{G}_r^{BB}$ to maximize $I(\boldsymbol{s},\tilde{\boldsymbol{y}}_d)$. According to [12], there exists an equivalent relationship between $I(\boldsymbol{s},\tilde{\boldsymbol{y}}_d)$ and the MSE matrix $\mathbf{E}_{MMSE}$.

$$I(\boldsymbol{s},\tilde{\boldsymbol{y}}_d) = \log_2\det(\mathbf{E}_{MMSE}^{-1}), \qquad (18)$$

where the MSE matrix $\mathbf{E}_{MMSE}$ is defined as the mean square error covariance matrix given the MMSE receiver.

The MMSE receiver is given as

$$\begin{aligned}\mathbf{W}_d^{MMSE} &= \operatorname{argmin}\mathbb{E}[\|\mathbf{W}_d^{BB}\tilde{\boldsymbol{y}}_d - \boldsymbol{s}\|^2] = (\tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB})^H \\ &\quad (\tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB}(\tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB})^H + \mathbf{R}_{\tilde{n}})^{-1},\end{aligned} \qquad (19)$$

where $\mathbf{R}_{\tilde{n}} = \sigma_1^2\tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\mathbf{W}_r^{RF}(\tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\mathbf{W}_r^{RF})^H + \sigma_2^2\mathbf{W}_r^{RF}(\mathbf{W}_r^{RF})^H$.

The MSE matrix $\mathbf{E}_{MMSE}$ is given as

$$\begin{aligned}\mathbf{E}_{MMSE} &= \mathbb{E}[(\mathbf{W}_d^{MMSE}\tilde{\boldsymbol{y}}_d - \boldsymbol{s})(\mathbf{W}_d^{MMSE}\tilde{\boldsymbol{y}}_d - \boldsymbol{s})^H] = (\mathbf{I}_{N_s} - \mathbf{W}_d^{MMSE}\tilde{\mathbf{H}}_2 \\ &\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB})(\mathbf{I}_{N_s} - \mathbf{W}_d^{MMSE}\tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB})^H + \mathbf{W}_d^{MMSE}\mathbf{R}_{\tilde{n}}(\mathbf{W}_d^{MMSE})^H\end{aligned} \qquad (20)$$

Substituting (19) into (20), we can express $\mathbf{E}_{MMSE}$ as

$$\mathbf{E}_{MMSE} = (\mathbf{I}_{N_s} + (\tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB})^H\mathbf{R}_{\tilde{n}}^{-1}\tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB})^{-1}. \qquad (21)$$

From (21), we can obtain (18).

Based on (18), we can establish the equivalence between the $I(\boldsymbol{s},\tilde{\boldsymbol{y}}_d)$ maximization problem and a WMMSE problem as [12] did.

The $I(\boldsymbol{s},\tilde{\boldsymbol{y}}_d)$ maximization problem is formulated as

$$\begin{aligned}&\min_{\mathbf{F}_t^{BB},\mathbf{G}_r^{BB}} -I(\boldsymbol{s},\tilde{\boldsymbol{y}}_d) \\ &s.t. \ \left\|\mathbf{F}_t^{RF}\mathbf{F}_t^{BB}\right\|_F^2 \leq E_t \\ &\quad \left\|\mathbf{F}_r^{RF}\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB}\boldsymbol{s} + \mathbf{F}_r^{RF}\mathbf{G}_r^{BB}\mathbf{W}_r^{RF}\boldsymbol{n}_1\right\|_F^2 \leq E_r.\end{aligned} \qquad (22)$$

The WMMSE problem is formulated as

$$\begin{aligned}&\min_{\mathbf{F}_t^{BB},\mathbf{G}_r^{BB},\mathbf{V}} \operatorname{Tr}(\mathbf{V}\mathbf{E}_{MMSE}) \\ &s.t. \ \left\|\mathbf{F}_t^{RF}\mathbf{F}_t^{BB}\right\|_F^2 \leq E_t \\ &\quad \left\|\mathbf{F}_r^{RF}\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB}\boldsymbol{s} + \mathbf{F}_r^{RF}\mathbf{G}_r^{BB}\mathbf{W}_r^{RF}\boldsymbol{n}_1\right\|_F^2 \leq E_r,\end{aligned} \qquad (23)$$

where $\mathbf{V}$ is a constant weight matrix.

We will show that Problems (22) and (23) have same optimum, i.e., the points that satisfy the KKT conditions for (22) and (23) are the same. Same as [12], we set the partial derivatives of the Lagrange functions of (22) and (23) to zero. Note that the power constraints of (22) and (23) are the same. To prove the equivalence, we only need to calculate the partial derivatives of $-I(\boldsymbol{s},\tilde{\boldsymbol{y}}_d)$ and $\operatorname{Tr}(\mathbf{V}\mathbf{E}_{MMSE})$ w.r.t $\mathbf{F}_t^{BB}$ and $\mathbf{G}_r^{BB}$, respectively. Note that $\partial\log\det(\mathbf{X}) = \operatorname{Tr}(\mathbf{X}^{-1}\partial\mathbf{X})$. Taking $\mathbf{F}_t^{BB}$ as an example, for $I(\boldsymbol{s},\tilde{\boldsymbol{y}}_d)$, we have

$$\frac{\partial - I(\boldsymbol{s},\tilde{\boldsymbol{y}}_d)}{\partial\mathbf{F}_t^{BB}} = -\frac{\partial\log_2\det(\mathbf{E}_{MMSE}^{-1})}{\partial\mathbf{F}_t^{BB}} = -\frac{\operatorname{Tr}(\mathbf{E}_{\mathbf{MMSE}}\partial\mathbf{E}_{MMSE}^{-1})}{\partial\mathbf{F}_t^{BB}}, \qquad (24)$$

Note that $\partial\mathbf{X}^{-1} = -\mathbf{X}^{-1}(\partial\mathbf{X})\mathbf{X}^{-1}$ and $\partial(\mathbf{A}\mathbf{X}) = \partial(\mathbf{X})\mathbf{A} + \partial(\mathbf{A})\mathbf{X}$. For $\operatorname{Tr}(\mathbf{V}\mathbf{E}_{MMSE})$, we have

$$\begin{aligned}\frac{\partial\operatorname{Tr}(\mathbf{V}\mathbf{E}_{MMSE})}{\partial\mathbf{F}_t^{BB}} &= -\frac{\operatorname{Tr}(\partial(\mathbf{V}(\mathbf{E}_{MMSE}^{-1})^{-1}))}{\partial\mathbf{F}_t^{BB}} \\ &= -\frac{\operatorname{Tr}(\mathbf{E}_{MMSE}\partial(\mathbf{E}_{MMSE}^{-1})\mathbf{E}_{MMSE}\mathbf{V} + \partial(\mathbf{V})\mathbf{E}_{MMSE})}{\partial\mathbf{F}_t^{BB}} \\ &= -\frac{\operatorname{Tr}(\mathbf{E}_{MMSE}\partial(\mathbf{E}_{MMSE}^{-1})\mathbf{E}_{MMSE}\mathbf{V})}{\partial\mathbf{F}_t^{BB}}.\end{aligned} \qquad (25)$$

If we set the constant weight matrix $\mathbf{V} = \mathbf{E}_{MMSE}^{-1}$, then we can have

$$\frac{\partial - I(\boldsymbol{s},\tilde{\boldsymbol{y}}_d)}{\partial\mathbf{F}_t^{BB}} = \frac{\partial\operatorname{Tr}(\mathbf{V}\mathbf{E}_{MMSE})}{\partial\mathbf{F}_t^{BB}}. \qquad (26)$$

Similarly, we can derive

$$\frac{\partial - I(\boldsymbol{s},\tilde{\boldsymbol{y}}_d)}{\partial\mathbf{G}_r^{BB}} = \frac{\partial\operatorname{Tr}(\mathbf{V}\mathbf{E}_{MMSE})}{\partial\mathbf{G}_r^{BB}}. \qquad (27)$$

From Eqs. (26) and (27), we can conclude that the KKT-conditions of (22) and (23) can be satisfied simultaneously, which suggests that it is possible to solve the mutual information maximization problem through the use of WMMSE by choosing an appropriate weight, i.e., $\mathbf{V}$. Therefore, we propose an iterative algorithm which deals with the WMMSE problem (23). The algorithm is described as follows:

1) Calculate the MMSE receiver $\mathbf{W}_d^{MMSE}$ in Eq. (19) and the MSE matrix $\mathbf{E}_{MMSE}$ in Eq. (20)
2) Update $\mathbf{V}$ by setting $\mathbf{V} = \mathbf{E}_{MMSE}^{-1}$
3) Fix $\mathbf{V}$ and $\mathbf{F}_t^{BB}$, then we find $\mathbf{G}_r^{BB}$ that minimizes $\operatorname{Tr}(\mathbf{V}\mathbf{E}_{MMSE}) = \operatorname{Tr}(\mathbf{V}((\mathbf{I}_{N_s} - \mathbf{W}_d^{MMSE}\tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB})(\mathbf{I}_{N_s} - \mathbf{W}_d^{MMSE}\tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB})^H + \mathbf{W}_d^{MMSE}\mathbf{R}_{\tilde{n}}(\mathbf{W}_d^{MMSE})^H))$ under the power constraints, i.e.,

$$\begin{aligned}\hat{\mathbf{G}}_r^{BB} &= \operatorname{argmin} \ \operatorname{Tr}(\mathbf{V}((\mathbf{I}_{N_s} - \mathbf{W}_d^{MMSE}\tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB}) \\ &\quad (\mathbf{I}_{N_s} - \mathbf{W}_d^{MMSE}\tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB})^H + \mathbf{W}_d^{MMSE}\mathbf{R}_{\tilde{n}}(\mathbf{W}_d^{MMSE})^H)) \\ &s.t. \ \left\|\mathbf{F}_r^{RF}\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB}\boldsymbol{s} + \mathbf{F}_r^{RF}\mathbf{G}_r^{BB}\mathbf{W}_r^{RF}\boldsymbol{n}_1\right\|_F^2 = E_r,\end{aligned} \qquad (28)$$

where we assume that the relay node uses the maximum available average transmit power, i.e., E_r, since the mutual information increases as the relay node power increases. Problem (28) is a convex optimization for $\mathbf{G}_r^{BB}$ and we can solve it using the KKT condition. Setting the partial derivative of the Lagrangian function of Problem (28) to zero, we can obtain the optimal solution for $\mathbf{G}_r^{BB}$, which is

$$\hat{\mathbf{G}}_r^{BB} = ((\mathbf{W}_d^{MMSE}\tilde{\mathbf{H}}_2)^H \mathbf{V}\mathbf{W}_d^{MMSE}\tilde{\mathbf{H}}_2 + \lambda^r (\mathbf{F}_r^{RF})^H \mathbf{F}_r^{RF})^{-1} (\tilde{\mathbf{H}}_1 \mathbf{F}_t^{BB} \mathbf{V}\tilde{\mathbf{H}}_2)^H (\tilde{\mathbf{H}}_1 \mathbf{F}_t^{BB} (\tilde{\mathbf{H}}_1 \mathbf{F}_t^{BB})^H + \sigma_1^2 \mathbf{W}_r^{RF} (\mathbf{W}_r^{RF})^H)^{-1}. \tag{29}$$

The Lagrange multiplier λ^r satisfies the power constraint and can be obtained by a bisection searching.

4) Fix $\mathbf{V}$ and $\mathbf{G}_r^{BB}$, then we find the $\mathbf{F}_t^{BB}$ to minimize $\mathrm{Tr}(\mathbf{V}\mathbf{E}_{MMSE}) = \mathrm{Tr}(\mathbf{V}((\mathbf{I}_{N_s} - \mathbf{W}_d^{MMSE}\tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB})(\mathbf{I}_{N_s} - \mathbf{W}_d^{MMSE}\tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB})^H + \mathbf{W}_d^{MMSE}\mathbf{R}_{\tilde{n}}(\mathbf{W}_d^{MMSE})^H))$ under the power constraints, i.e.,

$$\begin{aligned} \hat{\mathbf{F}}_t^{BB} = \arg\min \ & \mathrm{Tr}(\mathbf{V}((\mathbf{I}_{N_s} - \mathbf{W}_d^{MMSE}\tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB}) \\ & (\mathbf{I}_{N_s} - \mathbf{W}_d^{MMSE}\tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB})^H + \mathbf{W}_d^{MMSE}\mathbf{R}_{\tilde{n}}(\mathbf{W}_d^{MMSE})^H)) \\ s.t. \ & \left\|\mathbf{F}_t^{RF}\mathbf{F}_t^{BB}\right\|_F^2 = E_t \\ & \left\|\mathbf{F}_r^{RF}\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB}s + \mathbf{F}_r^{RF}\mathbf{G}_r^{BB}\mathbf{W}_r^{RF}\boldsymbol{n}_1\right\|_F^2 = E_r, \end{aligned} \tag{30}$$

where we assume that the source and relay nodes use the maximum available average transmit power, i.e., E_t and E_r, since the mutual information increases as the transmission power increases. Problem (30) is a convex optimization for $\mathbf{F}_t^{BB}$ and we can solve it using the KKT condition. The optimal solution for $\mathbf{F}_t^{BB}$ can be expressed as

$$\begin{aligned} \hat{\mathbf{F}}_t^{BB} = & ((\mathbf{W}_d^{MMSE}\tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1)^H \mathbf{V}\mathbf{W}_d^{MMSE}\tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1 \\ & + \lambda_1^t (\mathbf{F}_t^{RF})^H \mathbf{F}_t^{RF} + \lambda_2^t (\mathbf{F}_r^{RF}\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1)^H \mathbf{F}_r^{RF}\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1)^{-1} \\ & (\mathbf{V}\mathbf{W}_d^{MMSE}\tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1)^H, \end{aligned} \tag{31}$$

where λ_1^t and λ_2^t satisfy the power constraints. We can obtain λ_1^t and λ_2^t through a two-layer bisection search. The search algorithm is described in Alg. 1.

Note that the above algorithm converges; however, since the constant weight matrix $\mathbf{V}$ changes in each iteration, it does not generate a monotonic decreasing sequence. Therefore, we cannot directly prove the convergence of the proposed algorithm. Instead, using the same approach in [12], we can prove the convergence of the proposed algorithm by proving the monotonic convergence of an equivalent optimization problem. In other words, we can extend the convergence analysis procedure in [12] form multi-user MIMO scenarios to the relay scenarios. We skip the details of the convergence proof due to the lack of space.

D. $\mathbf{W}_d^{BB}$ design

Since $I(\boldsymbol{s}, \boldsymbol{y}_d) \le I(\boldsymbol{s}, \tilde{\boldsymbol{y}}_d)$ [11], after we get the maximum $I(\boldsymbol{s}, \tilde{\boldsymbol{y}}_d)$, the optimal $I(\boldsymbol{s}, \boldsymbol{y}_d)$ will be obtained if the destination node baseband processing does not cause any information loss. According to [13], MMSE-SIC is information lossless. Therefore, we use MMSE-SIC for our destination baseband design. To simplify the expression, let us define

Algorithm 1 Two-layer bisection search for λ_1^t and λ_2^t

```
initialize λ^t_{1,min} = λ^t_{2,min} = 0, λ^t_{1,max}, λ^t_{2,max};
while λ^t_{1,max} − λ^t_{1,min} > ε_1 do
    setting λ^t_1 = (λ^t_{1,min} + λ^t_{1,max}) / 2;
    while λ^t_{2,max} − λ^t_{2,min} > ε_2 do
        setting λ^t_2 = (λ^t_{2,min} + λ^t_{2,max}) / 2;
        calculate F^{BB}_t according to (31);
        if ||F^{RF}_r G^{BB}_r H̃_1 F^{BB}_t s + F^{RF}_r G^{BB}_r W^{RF}_r n_1||^2_F ≤ E_r then
            λ^t_{2,max} = λ^t_2;
        end if
        if ||F^{RF}_r G^{BB}_r H̃_1 F^{BB}_t s + F^{RF}_r G^{BB}_r W^{RF}_r n_1||^2_F ≥ E_r then
            λ^t_{2,min} = λ^t_2;
        end if
    end while
    calculate F^{BB}_t according to (31);
    if ||F^{RF}_t F^{BB}_t||^2_F ≤ E_t then
        λ^t_{1,max} = λ^t_1;
    end if
    if ||F^{RF}_t F^{BB}_t||^2_F ≥ E_t then
        λ^t_{1,min} = λ^t_1;
    end if
end while
```

$$\tilde{\boldsymbol{y}}_d = \tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB}\boldsymbol{s} + \tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\boldsymbol{n}}_1 + \tilde{\boldsymbol{n}}_2 = \mathbf{G}\boldsymbol{s} + \bar{\boldsymbol{v}}, \tag{32}$$

where $\mathbf{G} = [\boldsymbol{g}_1, ..., \boldsymbol{g}_{N_s}] \in \mathbf{C}^{N_{RF} \times N_s}$, $\bar{\boldsymbol{v}}$ is the colored noise with covariance matrix $\mathbf{R}_{\tilde{n}}$.

To implement the MMSE-SIC for the k^{th} stream, we subtract the effect of the first $k-1$ streams from the output and obtain

$$\tilde{\boldsymbol{y}}_d' = \tilde{\boldsymbol{y}}_d - \sum_{i=1}^{k-1} \boldsymbol{g}_i s_i + \bar{\boldsymbol{v}} = \boldsymbol{g}_k s_k + \sum_{j=k+1}^{N_s} \boldsymbol{g}_j s_j + \bar{\boldsymbol{v}} \tag{33}$$

Define $\mathbf{W}_d^{BB} = [\boldsymbol{w}_1, ..., \boldsymbol{w}_{N_s}]^H$, the baseband filter for the k^{th} stream is derived as the MMSE filter:

$$\boldsymbol{w}_k^H = \boldsymbol{g}_k^H (\sum_{j=k+1}^{N_s} \boldsymbol{g}_j \boldsymbol{g}_j^H + \mathbf{R}_{\tilde{n}})^{-1} \tag{34}$$

IV. Simulation Results

In the simulation, we consider a relay MIMO system consisting of one source node equipped with a $N_t = 64$ antenna array, a relay node with a $N_r = 32$ antenna array and a destination node with a $N_d = 48$ antenna array. The number of antennas is chosen from [10] for the purpose of the comparison. The channels are realized using Eq. (7). Due to the limited scattering characteristic of the mmWave channels, the number of paths should be less than the number of relay antennas. Here, we assume each channel has $L = 20$ paths. The φ_l of each path is assumed to be uniformly distributed in $[0, 2\pi]$. The results are averaged over 2000 channel realizations. The variances of AWGN noises σ_1 and σ_2 are assumed to be the same, i.e., $\sigma_1 = \sigma_2 = \sigma^2$.

In Fig. 3, we compare our algorithm with the ISA in [10] and the OMP in [9]. In the simulation, we assume the number of data streams is $N_s = 4$ and the number of RF chains is $N_{RF} = 6$. The fully-digital method is used as a benchmark, where we use the singular matrices of $\mathbf{H}_1$ and $\mathbf{H}_2$ as the precoding/combing matrices. We use the achievable data rate as the metric. Our algorithm outperforms the ISA by 4% and the OMP by 8.6% when SNR is 12 dB. Moreover, our algorithm has much lower complexity compared with the ISA in [10]. The ISA needs to solve three optimization sub-problems, and in each sub-problem it needs to solve an optimization problem through an iterative method. In our algorithm, we have closed-form solutions for each step. In addition, since we preform the baseband processing after the RF processing, the matrix dimensions are greatly reduced compared to [10]. Therefore, we have much faster convergence rate and lower complexity.

Fig. 4 shows the performance of our proposed algorithm in terms of the achievable rate and the performance of the ISA in terms of MSE with respect to the number of iterations. In addition to the fact that each iteration of our algorithm is less complex, our algorithm has a much faster convergence rate compared to the ISA.

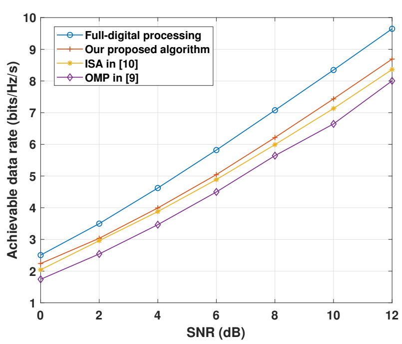

Fig. 3: Achievable rate comparison with $64 \times 32 \times 48$ when $N_s = 4$ and $N_{RF} = 6$

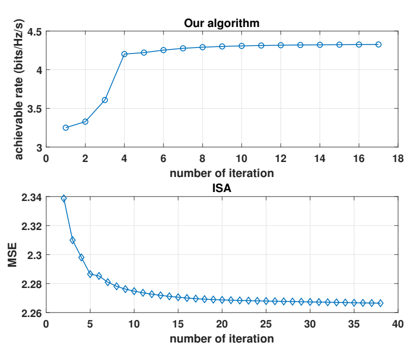

Fig. 4: Convergence rate comparison

V. Conclusion

In this paper, we considered mmWave AF relay networks in the domain of massive MIMO. We designed the hybrid precoding/combining matrices for the source node, the relay node, and the destination node. We first performed the RF processing to decompose the channel into parallel sub-channels by compensating the phase of each eigenmode of the channel. Given the RF processing matrices, we designed the baseband matrices to maximize the mutual information. The baseband processing is divided into two parts. We first jointly design the source node and the relay node by making use of the equivalence between maximizing $I(\boldsymbol{s}, \tilde{\boldsymbol{y}}_d)$ and the WMMSE. Given the optimal $\mathbf{F}_t^{BB}$ and $\mathbf{G}_r^{BB}$, we implement MMSE-SIC for $\mathbf{W}_d^{BB}$ to obtain the maximal $I(\boldsymbol{s}, \boldsymbol{y}_d)$. Simulation results show that our algorithm achieves better performance with lower complexity compared with other algorithms in the literature.

References

[1] T. S. Rappaport, J. R. W. Heath, R. C. Daniels, and J. N. Murdock, *Millimeter Wave Wireless Communications*. Pearson Education, 2014.

[2] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 101–107, June 2011.

[3] S. Han, I. Chih-Lin, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 186–194, Jan. 2015.

[4] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3180, Dec. 2004.

[5] Y. Jing and H. Jafarkhani, "Network beamforming using relays with perfect channel information," *IEEE Transactions on Information Theory*, vol. 55, no. 6, pp. 2499–2517, Jun. 2009.

[6] ——, "Single and multiple relay selection schemes and their achievable diversity orders," *IEEE Transactions on Wireless Communications*, vol. 8, no. 3, pp. 1414–1423, Mar. 2009.

[7] O. El Ayach, S. Rajagopal, S. AbuSurra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[8] X. Gao, L. Dai, and A. M. Sayeed, "Low RF-complexity technologies to enable millimeter-wave MIMO with large antenna array for 5G wireless communications," *IEEE Communications Magazine*, vol. 56, no. 4, pp. 211–217, Apr. 2018.

[9] J. Lee and Y. H. Lee, "AF relaying for millimeter wave communication systems with hybrid RF/baseband MIMO processing," in *proc of 2014 IEEE International Conference on Communications (ICC)*, Sydney, Australia.

[10] X. Xue, Y. Wang, L. Dai, and C. Masouros, "Relay hybrid precoding design in millimeter-wave massive MIMO systems," *IEEE Transactions on Signal Processing*, vol. 66, no. 8, pp. 2011–2026, Apr. 2018.

[11] R. M. Gray, *Entropy and information theory*. Springer Science & Business Media, 2011.

[12] S. S. Christensen, R. Agarwal, E. De Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.

[13] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge University Press, May 2005.

[14] W. Ni and X. Dong, "Hybrid block diagonalization for massive multiuser MIMO systems," *IEEE Transactions on Communications*, vol. 64, no. 1, pp. 201–211, Jan. 2016.

[15] T. S. Rappaport, F. Gutierrez, E. BenDor, J. N. Murdock, Y. Qiao, and J. I. Tamir, "Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 4, pp. 1850–1859, Apr. 2013.