MmWave Amplify-and-Forward MIMO Relay Networks with Hybrid Precoding/Combining Design

Lisi Jiang, 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 the 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. In addition, we also propose a robust joint transceiver design for imperfect channel state information.

I. INTRODUCTION

Communications over millimeter wave (mmWave) have received significant attention recently because of the high data rates provided by the large bandwidth at the mmWave carrier frequencies. Also, using large antenna arrays in mmWave communication systems is possible because

1

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 [2].

To overcome the large path losses, large antenna arrays can be placed at both transmitters and receivers to guarantee sufficient received signal power [3]. The large antenna arrays lead to a large number of radio frequency (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 [4].

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 [5]. Therefore, investigating the performance of hybrid precoding/combining in the relay scenario is important. For the conventional relay scenario, network beamforming in amplify-and-forward (AF) relay networks was studied in [6], [7].

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) [8], and (ii) sub-connected structure (where each RF chain is connected to a subset of antennas) [9]. For fully-connected mmWave networks with AF relay nodes, the authors in [10] designed hybrid precoding matrices using the orthogonal matching pursuit (OMP) algorithm. However, the performance of the OMP algorithm used in [10] depends on the orthogonality of the pre-determined candidates for the analog precoders. In [11], a joint source and relay precoding design for mmWave AF relay network is proposed based on semidefinite programming (SDP). However, the proposed method in [11] is only suitable for one data stream scenario. In [12], to reduce the complexity, the RF and the baseband (BB) are separately designed and a minimum mean squared error (MMSE)-based design for the BB filters is proposed. Although the algorithm in [12] shows its advantage over the OMP algorithm in terms of sum spectral efficiency, it did not optimize the sum rate of the system. In fact, [12] can be seen as a special case of our proposed methods since we minimize the weighted mean squared error. In [13], an efficient algorithm is proposed via employing the so-called Alternating Direction Method of Multipliers

(ADMM), which greatly reduces the distance between the hybrid precoder/combiner and the full-digital precoder/combiner. However, the ADMM algorithm has a high complexity and is sub-optimal in terms of the data rate for the system. For sub-connected structures, [14] proposes a MMSE-based relay hybrid precoding design. To make the problem tractable, [14] reformulates the original problem as three subproblems and proposes an iterative successive approximation (ISA) algorithm. The algorithm in [14] can also be extended to the fully-connected structure. Compared with the OMP algorithm, the ISA algorithm in [14] 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 multiple-input and multiple-output (MIMO) systems. To reduce the complexity, we separate the RF and the BB. For the RF, 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 by low-complexity methods. Based on the data processing inequality for the mutual information [15], we first jointly design the baseband source and relay nodes to maximize the mutual information before the destination baseband receiver. We propose a low-complexity iterative algorithm to design the precoder and combiner for the source and relay nodes, which is based on the equivalence between the mutual information maximization and the weighted MMSE [16]. After we obtain the optimal precoder and combiner for the source and relay nodes, we implement the MMSE successive interference cancellation (MMSE-SIC) filter [17] 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 [10]. Moreover, our algorithm achieves better performance with lower complexity compared to the ISA algorithm in [14].

We also propose a robust hybrid precoding/combining approach considering the inevitable imperfect channel state information (CSI) in the second part of the paper. Robust design for traditional relay systems has been well studied in papers, such as [18]–[20]. In [21], [22], the topic of imperfect channel state information in amplify-and-forward relay networks has been

studied under amplify-and-forward relay networks with limited feedback. However, there is not much work on the effects of imperfect channel state information in mmWave relay networks. In [23], a robust OMP-based algorithm is proposed to maximize the receiving signal-to-noise ratio (SNR) at the destination node. Similar with the non-robust case, the performance of the OMP-based algorithm depends on the orthogonality of the predetermined candidates for the analog precoders. In this paper, we adopt the well-known Kronecker model [18], [24] for the CSI mismatch. We first estimate the phase for RF precoding/combining to minimize the average estimation error. Then, we modify our proposed weighted MMSE approach for the perfect CSI to achieve a more robust performance for the baseband processing. Simulation results demonstrate the robustness of the proposed algorithm against CSI mismatch.

The contributions of our paper can be summarized as follows:

- We propose a hybrid precoding/combining approach for perfect CSI in mmWave relay systems. The phase-only RF precoding/combining matrices are first designed to achieve large array gains and support multi-stream transmissions. Then, we design the baseband processing system to achieve maximal mutual information by transforming the highly complicated non-convex mutual information maximization problem into an easily tractable weighted MMSE problem. An iterative algorithm which decouples the joint design into four sub-problems is developed.

- A robust design for the imperfect CSI is further proposed by modifying the non-robust precoding/combining design. To the best of our knowledge, except [23], this is the first mmWave relay system design that is robust to channel estimation error. Numerical results are provided to show the robustness of the proposed algorithm against CSI mismatch.

Compared with our conference version [1] which designs the hybrid filters for perfect CSI, we analyze the impact of imperfect CSI in this paper, and further propose a robust design to strengthen the robustness of our proposed algorithm. The remaining sections are organized as follows. In Section II , we describe the system model and the mmWave channel model. Section III formulates the proposed hybrid precoding/combing approach for the perfect CSI. Section IV presents the proposed robust hybrid design for the imperfect CSI. Numerical examples are presented and discussed in Section V. We provide concluding remarks in Section VI.

Notation: $\mathbb{C}^{m \times n}$ is the set of all $m \times n$ complex matrices with $\mathbb{C}^m \triangleq \mathbb{C}^{m \times 1}$ and $\mathbb{C} \triangleq \mathbb{C}^1$. $\mathbb{I}_m$ is the $m \times m$ identity matrix, and $\mathbf{0}_{m \times n}$ is the $m \times n$ all-zero matrix. $\mathbb{CN}(\boldsymbol{\mu}, \mathbf{K})$ is a circularly-symmetric complex Gaussian random vector with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{K}$.

Matrices $\mathbf{A}^T$ and $\mathbf{A}^H$ are the transpose and the Hermite transpose of matrix $\mathbf{A}$, respectively. Matrix $\mathbf{A} = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, ..., \boldsymbol{\alpha}_L]$ represents the concatenation of the L vectors $\boldsymbol{\alpha}_i$, and $\mathbf{B} = [\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_K]$ represents the concatenation of the K matrices $\mathbf{A}_i$.

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.
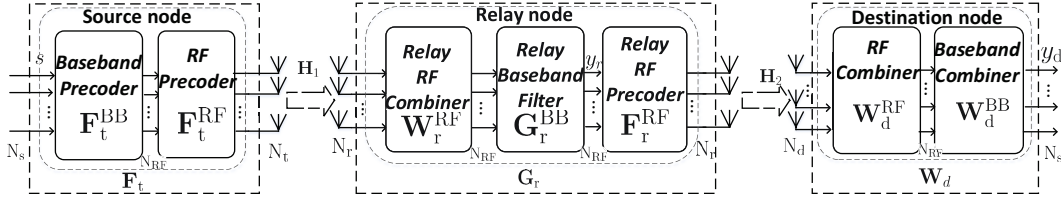

Fig. 1: System model

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 $\mathbf{H}_1 \in \mathbb{C}^{N_r \times N_t}$. The relay receives the signal with an RF combiner $\mathbf{W}_r^{RF} \in \mathbb{C}^{N_{RF} \times N_r}$ 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 $\mathbf{F}_r^{RF} \in \mathbb{C}^{N_r \times N_{RF}}$ through a MIMO channel $\mathbf{H}_2 \in \mathbb{C}^{N_d \times N_r}$ and the destination receives the data with one RF combiner $\mathbf{W}_d^{RF} \in \mathbb{C}^{N_{RF} \times N_d}$ and one baseband combiner $\mathbf{W}_d^{BB} \in \mathbb{C}^{N_s \times N_{RF}}$.

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

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

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. (3) 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

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

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

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

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 the optimal data rate. However, finding the global optima for this problem (maxmizing R while imposing constant-amplitude on the RF analog precoder/combiners) is non-convex and intractable. Separated RF and baseband processing designs, as [25] 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 [26]–[28], which means the assumption of a rich scattering environment becomes 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 [2], [27]–[29] with L scatters. Each scatter is assumed to contribute N_{cl} propagation paths to the channel matrix H. Then, the channel is given by

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{L N_{cl}}} \sum_{l=1}^{L} \sum_{n=1}^{N_{cl}} \alpha_{l,n} \mathbf{a}_r(\varphi_{l,n}^r, \theta_{l,n}^r) \mathbf{a}_t^H(\varphi_{l,n}^t, \theta_{l,n}^t), \tag{7}$$

where $\alpha_{l,n}$ is the complex gain of the nth path in the lth scatter with the distribution $\mathbb{CN}(0,1)$, $\varphi_{l,n}^r(\theta_{l,n}^r)$ and $\varphi_{l,n}^t(\theta_{l,n}^t)$ are the random azimuth and elevation angles of arrival and departure. $\mathbf{a}_r(\varphi_{l,n}^r, \theta_{l,n}^r)$ and $\mathbf{a}_t(\varphi_{l,n}^t, \theta_{l,n}^t)$ are the receiving and transmitting antenna array response vectors, respectively. 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 [30]:

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

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

Since the channel in mmWave systems has limited scattering, we can further simplify the channel by assuming each scatter only contributes one path to the channel matrix. Then, the channel can be further expressed as

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{L}} \sum_{l=1}^{L} \alpha_l \mathbf{a}_r(\varphi_l^r, \theta_l^r) \mathbf{a}_t^H(\varphi_l^t, \theta_l^t). \tag{9}$$

The matrix formulation can be expressed as

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{L}} \mathbf{A}_r diag(\boldsymbol{\alpha}) \mathbf{A}_t^H, \tag{10}$$

where $\mathbf{A}_r = [\mathbf{a}_r(\varphi_1^r, \theta_1^r), ..., \mathbf{a}_r(\varphi_L^r, \theta_L^r)]$ and $\mathbf{A}_t = [\mathbf{a}_t(\varphi_1^t, \theta_1^t), ..., \mathbf{a}_t(\varphi_L^t, \theta_L^t)]$ 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 transmits 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 ith 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 value decomposition (SVD) for the channel matrix.

$$\mathbf{H}_1 = \mathbf{U}_1 \Sigma_1 \mathbf{V}_1^H = \sum_{i=1}^{L} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^H, \tag{11}$$

where $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ are the ith vectors in matrices $\mathbf{U}_1$ and $\mathbf{V}_1$, respectively, which correspond to σ_i. The singular values σ_is are assumed to be in a descending order. L is the rank of the channel and is equal to the number of propagation paths for the mmWave scenario. Note that for mmWave systems, the channels usually have limited scattering characteristics, which means the number of propagation paths is far less than $\min(N_t, N_r)$. In such cases, the channel rank is equal to the number of propagation paths L. Eq. (11) indicates that channel $\mathbf{H}_1$ has L eigenmodes. We denote the ith eigenmode by $\boldsymbol{u}_i \boldsymbol{v}_i^H$, and its gain by σ_i.

For our RF precoding/combining, we want to maximize the projection of the ith data stream onto the ith eigenmode, i.e., $\left| \boldsymbol{w}_i^H \boldsymbol{u}_i \boldsymbol{v}_i^H \boldsymbol{f}_i \right|$, where $\boldsymbol{f}_i$ and $\boldsymbol{w}_i$ are the ith vector of precoder $\mathbf{F}_t^{RF}$ and combiner $\mathbf{W}_r^{RF}$, respectively. 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 ith data stream onto the ith eigenmode of the channel, will satisfy the following conditions:*

$$phase(\boldsymbol{f}_i[m]) = phase(\boldsymbol{v}_i[m]) \; \forall m = 1, 2, ..., N_t, \tag{12}$$

$$phase(\boldsymbol{w}_i[n]) = phase(\boldsymbol{u}_i[n]) \; \forall n = 1, 2, ..., N_r, \tag{13}$$

where $\cdot[k]$ represents the kth 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 $\boldsymbol{u}_i = [\rho_1^i e^{j\beta_1^i}, \rho_2^i e^{j\beta_2^i}, ..., \rho_{N_r}^i e^{j\beta_{N_r}^i}]^T$, respectively. Then, the projection can be calculated as

$$\left| \boldsymbol{w}_i^H \boldsymbol{u}_i \boldsymbol{v}_i^H \boldsymbol{f}_i \right| = \left| \frac{1}{\sqrt{N_r}} \sum_{n=1}^{N_r} \rho_n^i e^{j(\varphi_n^i - \beta_n^i)} \right| \left| \frac{1}{\sqrt{N_t}} \sum_{m=1}^{N_t} r_m^i e^{j(\alpha_m^i - \theta_m^i)} \right|. \tag{14}$$

According to the Cauchy-Schwartz inequality, we have

$$\left| \frac{1}{\sqrt{N_t}} \sum_{m=1}^{N_t} r_m^i e^{j(\alpha_m^i - \theta_m^i)} \right|^2$$
$$\leq \frac{1}{N_t} \sum_{m=1}^{N_t} \left| r_m^i \right|^2 \sum_{m=1}^{N_t} \left| e^{j(\alpha_m^i - \theta_m^i)} \right|^2 = \frac{1}{N_t} \sum_{m=1}^{N_t} \left| r_m^i \right|^2. \tag{15}$$

Equality can be achieved in (15) if and only if $\theta_m^i = \alpha_m^i \; \forall m = 1, 2, ..., N_t$. This means the maximal $\left| \boldsymbol{v}_i^H \boldsymbol{f}_i \right|$ is achieved when $\theta_m^i = \alpha_m^i \; \forall m = 1, 2, ..., N_t$. Similarly, the maximal $\left| \boldsymbol{w}_i^H \boldsymbol{u}_i \right|$ 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 $|\boldsymbol{w}_i^H \boldsymbol{u}_i \boldsymbol{v}_i^H \boldsymbol{f}_i|$, will satisfy the conditions in (12) and (13). □

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. $\boldsymbol{A}_t$ and $\boldsymbol{A}_r$ 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

$$\tilde{\mathbf{H}}_1 = \mathbf{W}_r^{RF}\mathbf{H}_1\mathbf{F}_t^{RF}, \tag{16}$$

$$\tilde{\mathbf{H}}_2 = \mathbf{W}_d^{RF}\mathbf{H}_2\mathbf{F}_r^{RF}. \tag{17}$$

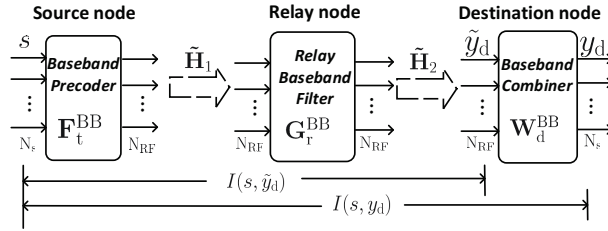Based on the equivalent channels, we simplify our system model as shown in Fig. 2.

Fig. 2: Baseband System model

Using the equivalent channels (16) and (17), we rewrite the received signals at the destination node as

$$\tilde{\boldsymbol{y}}_d = \tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB}\boldsymbol{s} + \tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\boldsymbol{n}}_1 + \tilde{\boldsymbol{n}}_2, \tag{18}$$

$$\boldsymbol{y}_d = \mathbf{W}_d^{BB}\tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB}\boldsymbol{s} + \mathbf{W}_d^{BB}\tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\boldsymbol{n}}_1 + \mathbf{W}_d^{BB}\tilde{\boldsymbol{n}}_2, \tag{19}$$

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 [15], $I(\boldsymbol{s}, \boldsymbol{y}_d) \leq I(\boldsymbol{s}, \tilde{\boldsymbol{y}}_d)$. 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 [17] 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 $\mathbf{F}_t^{\text{BB}}$ and $\mathbf{G}_r^{\text{BB}}$ to maximize $I(\boldsymbol{s}, \tilde{\boldsymbol{y}}_d)$. According to [16], there exists a relationship between $I(\boldsymbol{s}, \tilde{\boldsymbol{y}}_d)$ and the MSE matrix $\mathbf{E}_{\text{MMSE}}$, i.e.,

$$I(\boldsymbol{s}, \tilde{\boldsymbol{y}}_d) = \log_2 \det(\mathbf{E}_{\text{MMSE}}^{-1}), \tag{20}$$

where the MSE matrix $\mathbf{E}_{\text{MMSE}}$ is defined as the mean square error covariance matrix given the MMSE receiver. The detailed proof can be found in [16]. We give a brief derivation procedure below.

The MMSE receiver is defined as

$$\begin{aligned}
\mathbf{W}_d^{\text{MMSE}} = \arg\min \mathbb{E}[\|\mathbf{W}_d^{\text{BB}}\tilde{\boldsymbol{y}}_d - \boldsymbol{s}\|^2] = (\tilde{\mathbf{H}}_2 \mathbf{G}_r^{\text{BB}} \tilde{\mathbf{H}}_1 \mathbf{F}_t^{\text{BB}})^H \\
(\tilde{\mathbf{H}}_2 \mathbf{G}_r^{\text{BB}} \tilde{\mathbf{H}}_1 \mathbf{F}_t^{\text{BB}} (\tilde{\mathbf{H}}_2 \mathbf{G}_r^{\text{BB}} \tilde{\mathbf{H}}_1 \mathbf{F}_t^{\text{BB}})^H + \mathbf{R}_{\tilde{n}})^{-1},
\end{aligned} \tag{21}$$

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

The MMSE matrix $\mathbf{E}_{\text{MMSE}}$ can be calculated by

$$\begin{aligned}
\mathbf{E}_{\text{MMSE}} &= \mathbb{E}[(\mathbf{W}_d^{\text{MMSE}}\tilde{\boldsymbol{y}}_d - \boldsymbol{s})(\mathbf{W}_d^{\text{MMSE}}\tilde{\boldsymbol{y}}_d - \boldsymbol{s})^H] \\
&= (\mathbf{I}_{N_s} - \mathbf{W}_d^{\text{MMSE}} \tilde{\mathbf{H}}_2 \mathbf{G}_r^{\text{BB}} \tilde{\mathbf{H}}_1 \mathbf{F}_t^{\text{BB}}) \\
&\quad (\mathbf{I}_{N_s} - \mathbf{W}_d^{\text{MMSE}} \tilde{\mathbf{H}}_2 \mathbf{G}_r^{\text{BB}} \tilde{\mathbf{H}}_1 \mathbf{F}_t^{\text{BB}})^H + \mathbf{W}_d^{\text{MMSE}} \mathbf{R}_{\tilde{n}} (\mathbf{W}_d^{\text{MMSE}})^H.
\end{aligned} \tag{22}$$

Substituting (21) into (22), we can express $\mathbf{E}_{\text{MMSE}}$ as

$$\mathbf{E}_{\text{MMSE}} = (\mathbf{I}_{N_s} + (\tilde{\mathbf{H}}_2 \mathbf{G}_r^{\text{BB}} \tilde{\mathbf{H}}_1 \mathbf{F}_t^{\text{BB}})^H \mathbf{R}_{\tilde{n}}^{-1} \tilde{\mathbf{H}}_2 \mathbf{G}_r^{\text{BB}} \tilde{\mathbf{H}}_1 \mathbf{F}_t^{\text{BB}})^{-1}. \tag{23}$$

From (23), we can obtain (20).

Based on (20), we can establish the equivalence between the $I(\boldsymbol{s}, \tilde{\boldsymbol{y}}_d)$ maximization problem and a WMMSE problem as [16] did.

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

$$\begin{aligned}
&\min_{\mathbf{F}_t^{\text{BB}}, \mathbf{G}_r^{\text{BB}}} -I(\boldsymbol{s}, \tilde{\boldsymbol{y}}_d) \\
&s.t. \ \left\|\mathbf{F}_t^{\text{RF}} \mathbf{F}_t^{\text{BB}}\right\|_F^2 \leq \mathrm{E}_t, \\
&\qquad \mathbb{E}[\left\|\mathbf{F}_r^{\text{RF}} \mathbf{G}_r^{\text{BB}} \tilde{\mathbf{H}}_1 \mathbf{F}_t^{\text{BB}} \boldsymbol{s} + \mathbf{F}_r^{\text{RF}} \mathbf{G}_r^{\text{BB}} \mathbf{W}_r^{\text{RF}} \boldsymbol{n}_1\right\|_F^2] \leq \mathrm{E}_r.
\end{aligned} \tag{24}$$

The WMMSE problem is formulated as

$$\begin{aligned}
&\min_{\mathbf{F}_t^{\text{BB}}, \mathbf{G}_r^{\text{BB}}, \mathbf{V}} \mathrm{Tr}(\mathbf{V}\mathbf{E}_{\text{MMSE}}) \\
&s.t. \ \left\|\mathbf{F}_t^{\text{RF}} \mathbf{F}_t^{\text{BB}}\right\|_F^2 \leq \mathrm{E}_t, \\
&\qquad \mathbb{E}[\left\|\mathbf{F}_r^{\text{RF}} \mathbf{G}_r^{\text{BB}} \tilde{\mathbf{H}}_1 \mathbf{F}_t^{\text{BB}} \boldsymbol{s} + \mathbf{F}_r^{\text{RF}} \mathbf{G}_r^{\text{BB}} \mathbf{W}_r^{\text{RF}} \boldsymbol{n}_1\right\|_F^2] \leq \mathrm{E}_r,
\end{aligned} \tag{25}$$

11

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

We will show that Problems (24) and (25) have the same optimum solution, i.e., the points that satisfy the KKT conditions for (24) and (25) are the same. Same as [16], we set the partial derivatives of the Lagrange functions of (24) and (25) to zero. Note that the power constraints of (24) and (25) are the same. To prove the equivalence, we only need to calculate the partial derivatives of $-I(\boldsymbol{s}, \tilde{\mathbf{y}}_{\mathrm{d}})$ and $\mathrm{Tr}(\mathbf{V}\mathbf{E}_{\mathrm{MMSE}})$ w.r.t $\mathbf{F}_{\mathrm{t}}^{\mathrm{BB}}$ and $\mathbf{G}_{\mathrm{r}}^{\mathrm{BB}}$. Note that $\partial \log \det(\mathbf{X}) = \mathrm{Tr}(\mathbf{X}^{-1}\partial\mathbf{X})$. Taking $\mathbf{F}_{\mathrm{t}}^{\mathrm{BB}}$ as an example, for $I(\boldsymbol{s}, \tilde{\mathbf{y}}_{\mathrm{d}})$, we have

$$\frac{\partial -I(\boldsymbol{s}, \tilde{\mathbf{y}}_{\mathrm{d}})}{\partial \mathbf{F}_{\mathrm{t}}^{\mathrm{BB}}} = -\frac{\partial \log_2 \det(\mathbf{E}_{\mathrm{MMSE}})}{\partial \mathbf{F}_{\mathrm{t}}^{\mathrm{BB}}} = -\frac{\mathrm{Tr}(\mathbf{E}_{\mathrm{MMSE}}^{-1}\partial \mathbf{E}_{\mathrm{MMSE}})}{(\log 2)\partial \mathbf{F}_{\mathrm{t}}^{\mathrm{BB}}}, \tag{26}$$

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 $\mathrm{Tr}(\mathbf{V}\mathbf{E}_{\mathrm{MMSE}})$, we have

$$\begin{aligned}
\frac{\partial \mathrm{Tr}(\mathbf{V}\mathbf{E}_{\mathrm{MMSE}})}{\partial \mathbf{F}_{\mathrm{t}}^{\mathrm{BB}}} &= -\frac{\mathrm{Tr}(\partial(\mathbf{V}(\mathbf{E}_{\mathrm{MMSE}}^{-1})^{-1}))}{\partial \mathbf{F}_{\mathrm{t}}^{\mathrm{BB}}} \\
&= -\frac{\mathrm{Tr}(\mathbf{E}_{\mathrm{MMSE}}\partial(\mathbf{E}_{\mathrm{MMSE}}^{-1})\mathbf{E}_{\mathrm{MMSE}}\mathbf{V} + \partial(\mathbf{V})\mathbf{E}_{\mathrm{MMSE}})}{\partial \mathbf{F}_{\mathrm{t}}^{\mathrm{BB}}} \\
&= -\frac{\mathrm{Tr}(\mathbf{E}_{\mathrm{MMSE}}\partial(\mathbf{E}_{\mathrm{MMSE}}^{-1})\mathbf{E}_{\mathrm{MMSE}}\mathbf{V})}{\partial \mathbf{F}_{\mathrm{t}}^{\mathrm{BB}}}.
\end{aligned} \tag{27}$$

If we set the constant weight matrix $\mathbf{V} = \frac{\mathbf{E}_{\mathrm{MMSE}}^{-1}}{\log 2}$, then we have

$$\frac{\partial -I(\boldsymbol{s}, \tilde{\mathbf{y}}_{\mathrm{d}})}{\partial \mathbf{F}_{\mathrm{t}}^{\mathrm{BB}}} = \frac{\partial \mathrm{Tr}(\mathbf{V}\mathbf{E}_{\mathrm{MMSE}})}{\partial \mathbf{F}_{\mathrm{t}}^{\mathrm{BB}}}. \tag{28}$$

Similarly, we can derive

$$\frac{\partial -I(\boldsymbol{s}, \tilde{\mathbf{y}}_{\mathrm{d}})}{\partial \mathbf{G}_{\mathrm{r}}^{\mathrm{BB}}} = \frac{\partial \mathrm{Tr}(\mathbf{V}\mathbf{E}_{\mathrm{MMSE}})}{\partial \mathbf{G}_{\mathrm{r}}^{\mathrm{BB}}}. \tag{29}$$

From Eqs. (28) and (29), we can conclude that the KKT-conditions of (24) and (25) 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., $\mathbf{V}$. To maximize $I(\boldsymbol{s}, \tilde{\mathbf{y}}_d)$, we propose an iterative algorithm based on the WMMSE problem (25). Note that in the proposed algorithm, we also iteratively search for the appropriate weight matrix. When the algorithm converges, we will obtain the desired weight matrix as well as the optimal filters that maximize $I(\boldsymbol{s}, \tilde{\mathbf{y}}_{\mathrm{d}})$. The detailed algorithm is described as follows:

1) Calculate the MMSE receiver $\mathbf{W}_{\mathrm{d}}^{\mathrm{MMSE}}$ in Eq. (21) and the MSE matrix $\mathbf{E}_{\mathrm{MMSE}}$ in Eq. (22).
2) Update $\mathbf{V}$ by setting $\mathbf{V} = \frac{\mathbf{E}_{\mathrm{MMSE}}^{-1}}{\log 2}$.

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

$$\hat{\mathbf{G}}_r^{BB} = \operatorname{argmin}\ \mathrm{Tr}(\mathbf{V}((\mathbf{I}_{N_s} - \mathbf{W}_d^{\mathrm{MMSE}}\tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB})$$
$$(\mathbf{I}_{N_s} - \mathbf{W}_d^{\mathrm{MMSE}}\tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB})^H + \mathbf{W}_d^{\mathrm{MMSE}}\mathbf{R}_{\tilde{n}}(\mathbf{W}_d^{\mathrm{MMSE}})^H)) \tag{30}$$
$$s.t.\ \mathbb{E}[\|\mathbf{F}_r^{RF}\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB}\boldsymbol{s} + \mathbf{F}_r^{RF}\mathbf{G}_r^{BB}\mathbf{W}_r^{RF}\boldsymbol{n}_1\|_F^2] \leq \mathrm{E}_r.$$

Problem (30) is a convex optimization for $\mathbf{G}_r^{BB}$ and we can solve it using the KKT condition. Denoting the Lagrange function of Problem (30) as $L^r(\mathbf{G}_r^{BB}, \lambda^r) = \mathrm{Tr}(\mathbf{V}\mathbf{E}_{\mathrm{MMSE}}) + \lambda^r(\|\mathbf{F}_r^{RF}\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB}\boldsymbol{s} + \mathbf{F}_r^{RF}\mathbf{G}_r^{BB}\mathbf{W}_r^{RF}\boldsymbol{n}_1\|_F^2 - E_r)$, the KKT conditions are

$$\frac{\partial L^r(\mathbf{G}_r^{BB}, \lambda^r)}{\partial \mathbf{G}_r^{BB}} = 0, \tag{31}$$

$$\mathbb{E}[\|\mathbf{F}_r^{RF}\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB}\boldsymbol{s} + \mathbf{G}_r^{BB}\mathbf{W}_r^{RF}\boldsymbol{n}_1\|_F^2] - \mathrm{E}_r \leq 0, \tag{32}$$

$$\lambda^r(\mathbb{E}[\|\mathbf{F}_r^{RF}\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB}\boldsymbol{s} + \mathbf{G}_r^{BB}\mathbf{W}_r^{RF}\boldsymbol{n}_1\|_F^2] - \mathrm{E}_r) = 0, \tag{33}$$

$$\lambda^r \geq 0. \tag{34}$$

Solving (31), we have

$$\hat{\mathbf{G}}_r^{BB} = ((\mathbf{W}_d^{\mathrm{MMSE}}\tilde{\mathbf{H}}_2)^H \mathbf{V}\mathbf{W}_d^{\mathrm{MMSE}}\tilde{\mathbf{H}}_2 + \lambda^r(\mathbf{F}_r^{RF})^H\mathbf{F}_r^{RF})^{-1}$$
$$(\tilde{\mathbf{H}}_2)^H(\mathbf{W}_d^{\mathrm{MMSE}})^H\mathbf{V}(\mathbf{F}_t^{BB})^H(\tilde{\mathbf{H}}_1)^H \tag{35}$$
$$(\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB}(\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB})^H + \sigma_1^2\mathbf{W}_r^{RF}(\mathbf{W}_r^{RF})^H)^{-1}.$$

Based on (32) and (33), we can obtain the Lagrange multiplier λ^r as follows. First, we calculate $\hat{\mathbf{G}}_r^{BB}$ by setting $\lambda^r = 0$. If the power constraint is satisfied, then we set $\lambda^r = 0$. If the power constraint is not satisfied, then, we initialize λ^r with a pre-defined value and substitute it into (32) and start a bisection search for λ^r until the power constraint is satisfied.

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

$$\hat{\mathbf{F}}_t^{BB} = \operatorname{argmin}\ \mathrm{Tr}(\mathbf{V}((\mathbf{I}_{N_s} - \mathbf{W}_d^{\mathrm{MMSE}}\tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB})(\mathbf{I}_{N_s}$$
$$- \mathbf{W}_d^{\mathrm{MMSE}}\tilde{\mathbf{H}}_2\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB})^H + \mathbf{W}_d^{\mathrm{MMSE}}\mathbf{R}_{\tilde{n}}(\mathbf{W}_d^{\mathrm{MMSE}})^H))$$
$$s.t.\ \|\mathbf{F}_t^{RF}\mathbf{F}_t^{BB}\|_F^2 \leq \mathrm{E}_t, \tag{36}$$
$$\mathbb{E}[\|\mathbf{F}_r^{RF}\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB}s + \mathbf{F}_r^{RF}\mathbf{G}_r^{BB}\mathbf{W}_r^{RF}\boldsymbol{n}_1\|_F^2] \leq \mathrm{E}_r.$$

Problem (36) is a convex optimization for $\mathbf{F}_t^{BB}$ and, similar with Problem (30), we can solve Problem (36) using the KKT conditions. Denoting the Lagrange function of Problem (36) as $L^t(\mathbf{F}_t^{BB}, \lambda_1^t, \lambda_2^t)$, the KKT conditions are

$$\frac{\partial L^t(\mathbf{F}_t^{BB}, \lambda_1^t, \lambda_2^t)}{\partial \mathbf{F}_t^{BB}} = 0, \tag{37}$$

$$\left\| \mathbf{F}_t^{RF} \mathbf{F}_t^{BB} \right\|_F^2 - E_t \leq 0, \tag{38}$$

$$\mathbb{E}[\left\| \mathbf{F}_r^{RF} \mathbf{G}_r^{BB} \tilde{\mathbf{H}}_1 \mathbf{F}_t^{BB} s + \mathbf{F}_r^{RF} \mathbf{G}_r^{BB} \mathbf{W}_r^{RF} \boldsymbol{n}_1 \right\|_F^2] - E_r \leq 0, \tag{39}$$

$$\lambda_1^t(\left\| \mathbf{F}_t^{RF} \mathbf{F}_t^{BB} \right\|_F^2 - E_t) = 0, \tag{40}$$

$$\lambda_2^t(\mathbb{E}[\left\| \mathbf{F}_r^{RF} \mathbf{G}_r^{BB} \tilde{\mathbf{H}}_1 \mathbf{F}_t^{BB} s + \mathbf{F}_r^{RF} \mathbf{G}_r^{BB} \mathbf{W}_r^{RF} \boldsymbol{n}_1 \right\|_F^2] - E_r) = 0, \tag{41}$$

$$\lambda_1^t, \lambda_2^t \geq 0. \tag{42}$$

The optimal solution for $\mathbf{F}_t^{BB}$ can be expressed as

$$\begin{aligned} \hat{\mathbf{F}}_t^{BB} = &((\mathbf{W}_d^{MMSE} \tilde{\mathbf{H}}_2 \mathbf{G}_r^{BB} \tilde{\mathbf{H}}_1)^H \mathbf{V} \mathbf{W}_d^{MMSE} \tilde{\mathbf{H}}_2 \mathbf{G}_r^{BB} \tilde{\mathbf{H}}_1 \\ &+ \lambda_1^t (\mathbf{F}_t^{RF})^H \mathbf{F}_t^{RF} + \lambda_2^t (\mathbf{F}_r^{RF} \mathbf{G}_r^{BB} \tilde{\mathbf{H}}_1)^H \mathbf{F}_r^{RF} \mathbf{G}_r^{BB} \tilde{\mathbf{H}}_1)^{-1} \\ &(\mathbf{V} \mathbf{W}_d^{MMSE} \tilde{\mathbf{H}}_2 \mathbf{G}_r^{BB} \tilde{\mathbf{H}}_1)^H, \end{aligned} \tag{43}$$

where λ_1^t and λ_2^t are the non-negative Lagrange multipliers corresponding to the power constraints. Similar with Problem (30), to obtain λ_1^t and λ_2^t, we consider four cases: i) if the power constraints are satisfied when $\lambda_1^t = 0$ and $\lambda_2^t = 0$, we will set λ_1^t and λ_2^t equal to 0; ii) if case (i) is not satisfied, we then set $\lambda_1^t = 0$ and start a bisection search for λ_2^t until the KKT condition (41) and the power constraint (38) are satisfied; iii) if (41) and (38) cannot be satisfied simultaneously through the bisection search for λ_2^t, we then set $\lambda_2^t = 0$ and start a bisection search for λ_1^t until (39) and (40) are satisfied; iv) if we cannot find the appropriate λ_1^t to satisfy (39) and (40) at the same time, we can obtain λ_1^t and λ_2^t through a two-layer bisection search. The search algorithm is described in Algorithm 1.

The entire design procedures for $\mathbf{F}_t^{BB}$ and $\mathbf{G}_r^{BB}$ are summarized in Algorithm 2.

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

1: initialize $\lambda_{1,\min}^t = \lambda_{2,\min}^t = 0$, $\lambda_{1,\max}^t$, $\lambda_{2,\max}^t$;

2: **while** $\lambda_{1,\max}^t - \lambda_{1,\min}^t > \varepsilon_1$ **do**

3:     setting $\lambda_1^t = \frac{\lambda_{1,\min}^t + \lambda_{1,\max}^t}{2}$;

4:     **while** $\lambda_{2,\max}^t - \lambda_{2,\min}^t > \varepsilon_2$ **do**

5:         setting $\lambda_2^t = \frac{\lambda_{2,\min}^t + \lambda_{2,\max}^t}{2}$;

6:         calculate $\mathbf{F}_t^{BB}$ according to (43);

7:         **if** $\mathbb{E}[\|\mathbf{F}_r^{RF}\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB}\boldsymbol{s} + \mathbf{F}_r^{RF}\mathbf{G}_r^{BB}\mathbf{W}_r^{RF}\boldsymbol{n}_1\|_F^2] \leq \mathrm{E}_r$ **then**

8:             $\lambda_{2,\max}^t = \lambda_2^t$;

9:         **end if**

10:        **if** $\mathbb{E}[\|\mathbf{F}_r^{RF}\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB}\boldsymbol{s} + \mathbf{F}_r^{RF}\mathbf{G}_r^{BB}\mathbf{W}_r^{RF}\boldsymbol{n}_1\|_F^2] \geq \mathrm{E}_r$ **then**

11:            $\lambda_{2,\min}^t = \lambda_2^t$;

12:        **end if**

13:     **end while**

14:     calculate $\mathbf{F}_t^{BB}$ according to (43);

15:     **if** $\|\mathbf{F}_t^{RF}\mathbf{F}_t^{BB}\|_F^2 \leq \mathrm{E}_t$ **then**

16:         $\lambda_{1,\max}^t = \lambda_1^t$;

17:     **end if**

18:     **if** $\|\mathbf{F}_t^{RF}\mathbf{F}_t^{BB}\|_F^2 \geq \mathrm{E}_t$ **then**

19:         $\lambda_{1,\min}^t = \lambda_1^t$;

20:     **end if**

21: **end while**

D. Convergence analysis

Since the constant weight matrix $\mathbf{V}$ changes at each iteration, it does not generate a monotonic decreasing sequence, which means we cannot directly prove the convergence of the proposed algorithm. Fortunately, according to [16], the iteration procedure to maximize the mutual information through minimizing WMMSE is the same optimization procedure for an equivalent optimization problem as below

$$
\min_{\substack{\mathbf{F}_t^{BB}, \mathbf{G}_r^{BB}, \\ \mathbf{V}, \mathbf{W}_d^{BB}}} \mathrm{Tr}(\mathbf{V}\mathbf{E}_{MMSE}) - \log\det(\mathbf{V})
$$

$$
s.t. \quad \|\mathbf{F}_t^{RF}\mathbf{F}_t^{BB}\|_F^2 \leq \mathrm{E}_t,
$$

$$
\mathbb{E}[\|\mathbf{F}_r^{RF}\mathbf{G}_r^{BB}\tilde{\mathbf{H}}_1\mathbf{F}_t^{BB}\boldsymbol{s} + \mathbf{F}_r^{RF}\mathbf{G}_r^{BB}\mathbf{W}_r^{RF}\boldsymbol{n}_1\|_F^2] \leq \mathrm{E}_r.
$$

(44)

Algorithm 2 Design for $\mathbf{F}_t^{\text{BB}}$ and $\mathbf{G}_r^{\text{BB}}$

1: Initialize $\mathbf{F}_t^{\text{BB}(0)}$ and $\mathbf{G}_r^{\text{BB}(0)}$;

2: Set $k = 1$;

3: **while** $|I(\boldsymbol{s}, \tilde{\mathbf{y}}_d)^{(k)} - I(\boldsymbol{s}, \tilde{\mathbf{y}}_d)^{(k-1)}| > \varepsilon$ **do**

4:      Calculate the MMSE receiver $\mathbf{W}_d^{\text{MMSE}(k)}$ according to (21) and the MSE matrix $\mathbf{E}_{\text{MMSE}}^{(k)}$ in Eq. (22);

5:      Update $\mathbf{V}$ by setting $\mathbf{V} = \frac{(\mathbf{E}_{\text{MMSE}}^{(k)})^{-1}}{\log 2}$;

6:      Calculate $\mathbf{F}_t^{\text{BB}(k)}$ as Step III illustrates;

7:      Calculate $\mathbf{G}_r^{\text{BB}(k)}$ as Step IV illustrates;

8:      $k = k + 1$;

9: **end while**

The proof of the equivalence is similar to the proof in [16] and we omit the detailed prove for brevity. The main idea is that the alternating minimization of the objective in (44) corresponds to Steps 1-4 in our proposed algorithm. For example, when $\mathbf{F}_t^{\text{BB}}, \mathbf{G}_r^{\text{BB}}$ and $\mathbf{V}$ are fixed, optimizing (44) w.r.t. $\mathbf{W}_d^{\text{BB}}$ becomes minimizing MMSE, which gives the same result as Step 1. When $\mathbf{F}_t^{\text{BB}}, \mathbf{G}_r^{\text{BB}}$ and $\mathbf{W}_d^{\text{BB}}$ are fixed, the optimal solution for $\mathbf{V}$ which minimizes the objective $\text{Tr}(\mathbf{VE}_{\text{MMSE}}) - \log \det(\mathbf{V})$ in (44) is the same as Step 2.

Based on this equivalence, we can prove the convergence of the proposed algorithm by proving the monotonic convergence of Problem (44). According to [16], the objective in Problem (44) has a lower bound, which is the negative of the maximum mutual information. Due to the alternating minimization process, the objective in Problem (44) decreases monotonically. Since a sequence of monotonically decreasing numbers with a lower bound converges, Problem (44) converges and so does our proposed algorithm.

E. Complexity analysis

Since we provide closed-form solutions for each iteration, the main complexity lies in the search for the appropriate Lagrange multipliers. Let us define the search accuracy as ε. This is a relative measure for the search interval. For example, if the length of our search interval is l, then the threshold for the search termination is set to be εl. Based on the accuracy ε, the number of iterations in the bisection search in Step 3 is bounded by $\mathscr{O}(\log_2 \frac{1}{\varepsilon})$. In Step 4, we use a two-layer bisection search, whose number of iteration is bounded by $\mathscr{O}(\log_2^2 \frac{1}{\varepsilon})$. So, for each outer iteration, the total number of inner iterations is $\mathscr{O}(\log_2 \frac{1}{\varepsilon}) + \mathscr{O}(\log_2^2 \frac{1}{\varepsilon})$. Compared with the algorithm in [14], for which the number of inner iterations is $\mathscr{O}(2(2N_{\text{RF}}N_r)^{2.5} \log \frac{1}{\varepsilon})$ for

each outer iteration, the complexity of our algorithm is much lower especially for large antenna arrays.

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

Since $I(\boldsymbol{s}, \mathbf{y}_d) \leq I(\boldsymbol{s}, \tilde{\mathbf{y}}_d)$ [15], after we find the maximum $I(\boldsymbol{s}, \tilde{\mathbf{y}}_d)$, the optimal $I(\boldsymbol{s}, \mathbf{y}_d)$ will be obtained if the destination node baseband processing does not cause any information loss. According to [17], MMSE-SIC is information lossless. Therefore, we use MMSE-SIC for our destination baseband design. To simplify the expression, let us define

$$\tilde{\mathbf{y}}_d = \tilde{\mathbf{H}}_2 \mathbf{G}_r^{BB} \tilde{\mathbf{H}}_1 \mathbf{F}_t^{BB} \boldsymbol{s} + \tilde{\mathbf{H}}_2 \mathbf{G}_r^{BB} \tilde{\boldsymbol{n}}_1 + \tilde{\boldsymbol{n}}_2 = \mathbf{G} \boldsymbol{s} + \bar{\boldsymbol{v}}, \tag{45}$$

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 kth stream, we subtract the effect of the first $k-1$ streams from the output and obtain

$$\tilde{\mathbf{y}}_{d'} = \tilde{\mathbf{y}}_d - \sum_{i=1}^{k-1} \boldsymbol{g}_i s_i + \bar{\boldsymbol{v}} = \boldsymbol{g}_k s_k + \sum_{j=k+1}^{N_s} \boldsymbol{g}_j s_j + \bar{\boldsymbol{v}}. \tag{46}$$

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

$$\boldsymbol{w}_k^H = \boldsymbol{g}_k^H \big(\sum_{j=k+1}^{N_s} \boldsymbol{g}_j \boldsymbol{g}_j^H + \mathbf{R}_{\tilde{n}} \big)^{-1}. \tag{47}$$

IV. ROBUST DESIGN

So far, we have designed the mmWave relay precoders/combiners under the perfect channel information. However, it is hard to avoid estimation/quantization errors while obtaining the channel information. To study the effects of imperfect channel estimation, we adopt the model provided in [18], [20], [24]. In this model, the relationship between the channel values and the corresponding estimated channel values are:

$$\mathbf{H}_1 = \bar{\mathbf{H}}_1 + \Phi_1^{\frac{1}{2}} \Delta_1 \Theta_1^{\frac{1}{2}}, \tag{48}$$

$$\mathbf{H}_2 = \bar{\mathbf{H}}_2 + \Phi_2^{\frac{1}{2}} \Delta_2 \Theta_2^{\frac{1}{2}}, \tag{49}$$

where $\mathbf{H}_1$ and $\mathbf{H}_2$ are the actual channel matrices, i.e., the channels that we cannot precisely estimate, and $\bar{\mathbf{H}}_1$ and $\bar{\mathbf{H}}_2$ are the estimated channels. The transmitting covariance matrix of the channel estimation error at the source node and the relay node are denoted by $\Theta_1 \in \mathbb{C}^{N_t \times N_t}$ and $\Theta_2 \in \mathbb{C}^{N_r \times N_r}$, respectively. The receiving covariance matrix of the channel estimation error at the

relay node and the destination node are denoted by $\mathbf{\Phi}_1 \in \mathbb{C}^{N_r \times N_r}$ and $\mathbf{\Phi}_2 \in \mathbb{C}^{N_d \times N_d}$, respectively. Δ_1 and Δ_2 are Gaussian random matrices with independent and identically distributed (i.i.d.) zero mean and unit variance entries and are the unknown parts of the CSI mismatch. we adopt the exponential model [18], [20] for the channel estimation error covariance matrices $\mathbf{\Phi}_1$, $\mathbf{\Theta}_1$, $\mathbf{\Phi}_2$ and $\mathbf{\Theta}_2$. To be specific, the entries of the matrices are given as $\mathbf{\Phi}_1(i,j) = \sigma_{e,1}^2 \beta_1^{|i-j|}$, $\mathbf{\Theta}_1(i,j) = \alpha_1^{|i-j|}$, $\mathbf{\Phi}_2(i,j) = \sigma_{e,2}^2 \beta_2^{|i-j|}$ and $\mathbf{\Theta}_2(i,j) = \alpha_2^{|i-j|}$, where α_1, β_1, α_2 and β_2 are the correlation coefficients and $\sigma_{e,1}^2$ and $\sigma_{e,2}^2$ denote the estimation error covariance. For simplicity, we assume $\alpha_1 = \alpha_2 = \alpha$, $\beta_1 = \beta_2 = \beta$ and $\sigma_{e,1}^2 = \sigma_{e,2}^2 = \sigma_e^2$.

As shown in Section V, the imperfect channel information will result in severe performance degradation. For example, the achievable data rate of [10], [14] can be decreased to half of what it is for the perfect CSI.

To further increase the robustness of our proposed algorithm, in this section, we will propose a robust precoding/combining design for the mmWave relay system based on our proposed WMMSE algorithm.

A. RF design

Recall that our RF precoding/combining is actually phase compensation for each eigenmode. The eigenmodes are obtained through SVD. When considering the imperfect CSI, the phase of each eigenmode cannot be precisely acquired. Let us take the actual channel $\mathbf{H}_1$ and the estimated $\bar{\mathbf{H}}_1$ as an example. The left singular matrices of $\mathbf{H}_1$ and $\bar{\mathbf{H}}_1$ are denoted by $\mathbf{U}_1$ and $\bar{\mathbf{U}}_1$. We denote the phase of each entry in $\mathbf{U}_1$ and $\bar{\mathbf{U}}_1$ by $\theta_{i,j}$ and $\bar{\theta}_{i,j}$, respectively. The phase difference in each entry can be calculated as $\Delta\theta_{i,j} = \theta_{i,j} - \bar{\theta}_{i,j}$. Let us assume that $\Delta\theta_{i,j}$ has a distribution $p(\Delta\theta)$. We want to make an estimation on $\Delta\theta_{i,j}$ to minimize the mean square estimation error $\mathbb{E}[(\Delta\hat{\theta}_{i,j} - \Delta\theta_{i,j})^2]$. The estimation $\Delta\hat{\theta}_{i,j} = \mathbb{E}[\Delta\theta_{i,j}]$ can be calculated based on the distribution $p(\Delta\theta)$. Note that we can only obtain the estimated channel. Once we calculate $\Delta\hat{\theta}_{i,j}$, we can calibrate the phase of each entry in $\bar{\mathbf{U}}_1$ as $\hat{\theta}_{i,j} = \bar{\theta}_{i,j} + \Delta\hat{\theta}_{i,j}$. Using the same approach, we can calibrate the phase of singular matrices of $\bar{\mathbf{H}}_1$ and $\bar{\mathbf{H}}_2$. Then, based on (12) and (13), we can calculate the RF precoders and combiners $\bar{\mathbf{F}}_t^{\mathrm{RF}}$, $\bar{\mathbf{W}}_r^{\mathrm{RF}}$, $\bar{\mathbf{F}}_r^{\mathrm{RF}}$ and $\bar{\mathbf{W}}_d^{\mathrm{RF}}$ based on the calibrated singular matrices of $\bar{\mathbf{H}}_1$ and $\bar{\mathbf{H}}_2$.

As we analyzed above, to calculate the RF precoders and combiners, we must know the distribution of $\Delta\theta_{i,j}$ to make the estimation $\Delta\hat{\theta}_{i,j} = \mathbb{E}[\Delta\theta_{i,j}]$. However, the theoretical analysis for the phase distribution is intractable. To obtain the phase distribution, we simulate 100 channel

18

realizations based on the imperfect channel models in (48), where we set $N_r = 32$, $N_t = 48$ and $L = 20$. We adopt the correlation model from [18], [20], [24] where the entries of the correlation matrices are selected as $\Phi_1(i,j) = \sigma_{e,1}^2 \beta_1^{|i-j|}$, $\Theta_1(i,j) = \alpha_1^{|i-j|}$. In the simulation, we set $\alpha_1 = 0$, $\beta_1 = 0$ and $\sigma_{e,1}^2 = 0.1$.

We collect the phase difference in each matrix entry from 60 simulations. In Fig. 3, we plot the simulated probability density function (PDF) of the $\Delta\theta$ in solid line. We use a generalized normal distribution [31] to approximate the distribution. The PDF of a generalized normal distribution is expressed as $f(x) = \frac{\beta}{2\alpha\Gamma(\frac{1}{\beta})} e^{-(|x-\mu|/\alpha)^\beta}$. We can see the approaching effect of different value of the shaping parameter β in Fig. 3. We use Kullback-Leibler distance as a performance measure for the approximation, which is calculated by $D_{\mathrm{KL}}(\boldsymbol{Y}||\boldsymbol{X}) = \sum_{i=1}^N \log(\frac{Y_i}{X_i})Y_i$ where $\boldsymbol{Y}$ and $\boldsymbol{X}$ are the probability distributions. The lower the Kullback-Leibler distance, the closer the two distributions are. Note that the absolute value of KL-distance varies when the number of total points (i.e., N) changes. In our simulation, the best approximation comes with the one with $\beta = 2$, since it has the lowest Kullback-Leibler distance. When $\beta = 2$, the generalized normal distribution in Fig. 3 is a Gaussian distribution with 0 mean, which means we can estimate $\Delta\hat{\theta}_{i,j} = 0$.
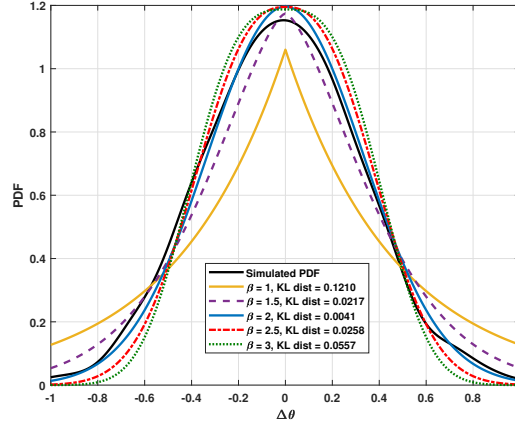

Fig. 3: Approximation of the Simulated PDF

B. Baseband design

Based on the RF precoders and combiners $\bar{\mathbf{F}}_t^{\mathrm{RF}}$, $\bar{\mathbf{W}}_r^{\mathrm{RF}}$, $\bar{\mathbf{F}}_r^{\mathrm{RF}}$ and $\bar{\mathbf{W}}_d^{\mathrm{RF}}$ designed in the last subsection, the equivalent baseband channels after the RF processing are

$$\begin{aligned}
\tilde{\mathbf{H}}_1 &= \bar{\mathbf{W}}_r^{\mathrm{RF}} \mathbf{H}_1 \bar{\mathbf{F}}_t^{\mathrm{RF}} = \bar{\mathbf{W}}_r^{\mathrm{RF}} \bar{\mathbf{H}}_1 \bar{\mathbf{F}}_t^{\mathrm{RF}} + \bar{\mathbf{W}}_r^{\mathrm{RF}} \Phi_1^{\frac{1}{2}} \Delta_1 \Theta_1^{\frac{1}{2}} \bar{\mathbf{F}}_t^{\mathrm{RF}} \\
&= \tilde{\bar{\mathbf{H}}}_1 + \tilde{\Phi}_1^{\frac{1}{2}} \Delta_1 \tilde{\Theta}_1^{\frac{1}{2}},
\end{aligned}$$

(50)

19

$$\tilde{\mathbf{H}}_2 = \bar{\mathbf{W}}_d^{RF}\mathbf{H}_2\bar{\mathbf{F}}_r^{RF} = \bar{\mathbf{W}}_d^{RF}\bar{\mathbf{H}}_2\bar{\mathbf{F}}_r^{RF} + \bar{\mathbf{W}}_d^{RF}\Phi_2^{\frac{1}{2}}\Delta_2\Theta_2^{\frac{1}{2}}\bar{\mathbf{F}}_r^{RF}$$
$$= \tilde{\bar{\mathbf{H}}}_2 + \check{\Phi}_2^{\frac{1}{2}}\Delta_2\tilde{\Theta}_2^{\frac{1}{2}}, \tag{51}$$

where we denote the true equivalent baseband channels by $\tilde{\mathbf{H}}_1$ and $\tilde{\mathbf{H}}_2$. The estimated equivalent baseband channels are denoted by $\tilde{\bar{\mathbf{H}}}_1$ and $\tilde{\bar{\mathbf{H}}}_2$, which are the channels we obtain at the source node. We define $\check{\Phi}_1 := \bar{\mathbf{W}}_r^{RF}\Phi_1(\bar{\mathbf{W}}_r^{RF})^H$, $\check{\Phi}_2 := \bar{\mathbf{W}}_d^{RF}\Phi_2(\bar{\mathbf{W}}_d^{RF})^H$, $\tilde{\Theta}_1 := (\bar{\mathbf{F}}_t^{RF})^H\Theta_1\bar{\mathbf{F}}_t^{RF}$ and $\tilde{\Theta}_2 := (\bar{\mathbf{F}}_r^{RF})^H\Theta_2\bar{\mathbf{F}}_r^{RF}$.

For (50) and (51) , we have the following properties [32] (Using $\tilde{\mathbf{H}}_1$ as an example)

$$\mathbb{E}_{\Delta_1}[\tilde{\mathbf{H}}_1\mathbf{C}\tilde{\mathbf{H}}_1^H] = \tilde{\bar{\mathbf{H}}}_1\mathbf{C}\tilde{\bar{\mathbf{H}}}_1^H + \mathrm{Tr}(\mathbf{C}\tilde{\Theta}_1)\check{\Phi}_1, \tag{52}$$

$$\mathbb{E}_{\Delta_1}[\tilde{\mathbf{H}}_1^H\mathbf{C}\tilde{\mathbf{H}}_1] = \tilde{\bar{\mathbf{H}}}_1^H\mathbf{C}\tilde{\bar{\mathbf{H}}}_1 + \mathrm{Tr}(\check{\Phi}_1\mathbf{C})\tilde{\Theta}_1. \tag{53}$$

To design a robust baseband system, we need to redesign the algorithm in Section III based on the imperfect channel models (50) and (51). The main idea is similar, i.e., that we first optimize the baseband filters $\bar{\mathbf{F}}_t^{BB}$ and $\bar{\mathbf{G}}_r^{BB}$ to maximize the average $\mathbb{E}_{\Delta_1,\Delta_2}[I(\boldsymbol{s},\tilde{\mathbf{y}}_d)]$, and then we use MMSE-SIC for $\bar{\mathbf{W}}_d^{BB}$. Note that we denote the baseband precoder/combiner based on the estimated equivalent baseband channels $\tilde{\bar{\mathbf{H}}}_1$ and $\tilde{\bar{\mathbf{H}}}_2$ by $\bar{\mathbf{F}}_t^{BB}$, $\bar{\mathbf{G}}_r^{BB}$ and $\bar{\mathbf{W}}_d^{BB}$. The main challenge here is if there still exists an equivalent relationship between the average mutual information maximization and the WMMSE minimization.

To derive the equivalent relationship, we first derive an upper bound for the average mutual information $\mathbb{E}_{\Delta_1,\Delta_2}[I(\boldsymbol{s},\tilde{\mathbf{y}}_d)]$ as

$$\mathbb{E}_{\Delta_1,\Delta_2}^{UB}[I(\boldsymbol{s},\tilde{\mathbf{y}}_d)] = \log_2\det(\mathbb{E}_{\Delta_1,\Delta_2}[\mathbf{I}_{N_s} + (\tilde{\mathbf{H}}_2\bar{\mathbf{G}}_r^{BB}\tilde{\mathbf{H}}_1\bar{\mathbf{F}}_t^{BB})^H\mathbf{R}_{\tilde{n}}^{-1}$$
$$\tilde{\mathbf{H}}_2\bar{\mathbf{G}}_r^{BB}\tilde{\mathbf{H}}_1\bar{\mathbf{F}}_t^{BB}]) = \log_2\det(\mathbf{I}_{N_s} + (\tilde{\bar{\mathbf{H}}}_2\bar{\mathbf{G}}_r^{BB}\tilde{\bar{\mathbf{H}}}_1\bar{\mathbf{F}}_t^{BB})^H\mathbf{R}_{\tilde{\bar{n}}}^{-1} \tag{54}$$
$$\tilde{\bar{\mathbf{H}}}_2\bar{\mathbf{G}}_r^{BB}\tilde{\bar{\mathbf{H}}}_1\bar{\mathbf{F}}_t^{BB} + \mathbf{B}_1 + \mathbf{B}_2),$$

where

$$\mathbf{R}_{\tilde{\bar{n}}} := \sigma_2^2\bar{\mathbf{W}}_r^{RF}(\bar{\mathbf{W}}_r^{RF})^H + \sigma_1^2(\tilde{\bar{\mathbf{H}}}_2\bar{\mathbf{G}}_r^{BB}\bar{\mathbf{W}}_r^{RF}(\tilde{\bar{\mathbf{H}}}_2\bar{\mathbf{G}}_r^{BB}\bar{\mathbf{W}}_r^{RF})^H$$
$$+ \mathrm{Tr}(\bar{\mathbf{G}}_r^{BB}\bar{\mathbf{W}}_r^{RF}(\bar{\mathbf{G}}_r^{BB}\bar{\mathbf{W}}_r^{RF})^H)\tilde{\Theta}_2)\check{\Phi}_2),$$

$$\mathbf{B}_1 := (\bar{\mathbf{G}}_r^{BB}\tilde{\bar{\mathbf{H}}}_1\bar{\mathbf{F}}_t^{BB})^H\mathrm{Tr}(\check{\Phi}_2\mathbf{R}_{\tilde{\bar{n}}})\tilde{\Theta}_2\bar{\mathbf{G}}_r^{BB}\tilde{\bar{\mathbf{H}}}_1\bar{\mathbf{F}}_t^{BB},$$

$$\mathbf{B}_2 := (\bar{\mathbf{F}}_t^{BB})^H\mathrm{Tr}(\check{\Phi}_1((\tilde{\bar{\mathbf{H}}}_2\bar{\mathbf{G}}_r^{BB})^H\mathbf{R}_{\tilde{\bar{n}}}\tilde{\bar{\mathbf{H}}}_2\bar{\mathbf{G}}_r^{BB} +$$
$$(\bar{\mathbf{G}}_r^{BB})^H\mathrm{Tr}(\check{\Phi}_2\mathbf{R}_{\tilde{\bar{n}}})\tilde{\Theta}_2)\bar{\mathbf{G}}_r^{BB})\tilde{\Theta}_1\bar{\mathbf{F}}_t^{BB}.$$

20

Then, we need to derive the expression of $\mathbb{E}_{\Delta_1,\Delta_2}[\mathbf{E}_{\text{MMSE}}]$. According to (23), we have

$$
\begin{aligned}
&\mathbb{E}_{\Delta_1,\Delta_2}[\mathbf{E}_{\text{MMSE}}]\\
&= \mathbb{E}_{\Delta_1,\Delta_2}[(\mathbf{I}_{N_s} + (\tilde{\mathbf{H}}_2 \bar{\mathbf{G}}_r^{\text{BB}} \tilde{\mathbf{H}}_1 \bar{\mathbf{F}}_t^{\text{BB}})^H \mathbf{R}_{\tilde{n}}^{-1} \tilde{\mathbf{H}}_2 \bar{\mathbf{G}}_r^{\text{BB}} \tilde{\mathbf{H}}_1 \bar{\mathbf{F}}_t^{\text{BB}})^{-1}]\\
&= (\mathbf{I}_{N_s} + (\tilde{\tilde{\mathbf{H}}}_2 \bar{\mathbf{G}}_r^{\text{BB}} \tilde{\tilde{\mathbf{H}}}_1 \bar{\mathbf{F}}_t^{\text{BB}})^H \mathbf{R}_{\tilde{\tilde{n}}}^{-1} \tilde{\tilde{\mathbf{H}}}_2 \bar{\mathbf{G}}_r^{\text{BB}} \tilde{\tilde{\mathbf{H}}}_1 \bar{\mathbf{F}}_t^{\text{BB}} + \mathbf{B}_1 + \mathbf{B}_2)^{-1}.
\end{aligned}
\tag{55}
$$

Equation (55) implies that the relationship $\mathbb{E}_{\Delta_1,\Delta_2}^{\text{UB}}[I(\boldsymbol{s},\tilde{\boldsymbol{y}}_d)] = \log_2 \det(\mathbb{E}_{\Delta_1,\Delta_2}[\mathbf{E}_{\text{MMSE}}]^{-1})$ still holds for the imperfect channel model, which means we can maximize the upper bound of the average mutual information through the WMMSE minimization as discussed in Section III.

The expression of $\mathbb{E}_{\Delta_1,\Delta_2}[\mathbf{E}_{\text{MMSE}}]$ in (55) includes a matrix inverse operator, which complicates the following calculation for $\bar{\mathbf{F}}_t^{\text{BB}}$ and $\bar{\mathbf{G}}_r^{\text{BB}}$. To derive a simpler expression for $\mathbb{E}_{\Delta_1,\Delta_2}[\mathbf{E}_{\text{MMSE}}]$, we first calculate the average MSE matrix $\mathbb{E}_{\Delta_1,\Delta_2}[\mathbf{E}_{\text{MSE}}]$. The MSE matrix is given by

$$
\begin{aligned}
\mathbb{E}_{\Delta_1,\Delta_2}[\mathbf{E}_{\text{MSE}}] &= \bar{\mathbf{W}}_d^{\text{BB}}(\mathbf{A} + \mathbf{R}_{\tilde{n}})(\bar{\mathbf{W}}_d^{\text{BB}})^H - (\bar{\mathbf{W}}_d^{\text{BB}})^H \tilde{\tilde{\mathbf{H}}}_2 \bar{\mathbf{G}}_r^{\text{BB}} \tilde{\tilde{\mathbf{H}}}_1 \bar{\mathbf{F}}_t^{\text{BB}}\\
&\quad - (\tilde{\tilde{\mathbf{H}}}_2 \bar{\mathbf{G}}_r^{\text{BB}} \tilde{\tilde{\mathbf{H}}}_1 \bar{\mathbf{F}}_t^{\text{BB}})^H \bar{\mathbf{W}}_d^{\text{BB}} + \mathbf{I}_{N_s},
\end{aligned}
\tag{56}
$$

where

$$
\mathbf{A} := \tilde{\tilde{\mathbf{H}}}_2 \bar{\mathbf{G}}_r^{\text{BB}} \mathbf{A}_1 (\tilde{\tilde{\mathbf{H}}}_2 \bar{\mathbf{G}}_r^{\text{BB}})^H + \text{Tr}(\bar{\mathbf{G}}_r^{\text{BB}} \mathbf{A}_1 (\bar{\mathbf{G}}_r^{\text{BB}})^H \tilde{\Theta}_2^H) \check{\Phi}_2,
$$

$$
\mathbf{A}_1 := \tilde{\tilde{\mathbf{H}}}_1 \bar{\mathbf{F}}_t^{\text{BB}} (\tilde{\tilde{\mathbf{H}}}_1 \bar{\mathbf{F}}_t^{\text{BB}})^H + \text{Tr}(\bar{\mathbf{F}}_t^{\text{BB}} (\bar{\mathbf{F}}_t^{\text{BB}})^H \tilde{\Theta}_1^H) \check{\Phi}_1.
$$

Based on (56), we can derive the $\bar{\mathbf{W}}_d^{\text{MMSE}}$, which minimizes $\mathbb{E}_{\Delta_1,\Delta_2}[\mathbf{E}_{\text{MSE}}]$, as

$$
\bar{\mathbf{W}}_d^{\text{MMSE}} = (\tilde{\tilde{\mathbf{H}}}_2 \bar{\mathbf{G}}_r^{\text{BB}} \tilde{\tilde{\mathbf{H}}}_1 \bar{\mathbf{F}}_t^{\text{BB}})^H (\mathbf{A} + \mathbf{R}_{\tilde{n}})^{-1}.
\tag{57}
$$

Substituting (57) into (56), we have

$$
\begin{aligned}
\mathbb{E}_{\Delta_1,\Delta_2}[\mathbf{E}_{\text{MMSE}}] &= \bar{\mathbf{W}}_d^{\text{MMSE}}(\mathbf{A} + \mathbf{R}_{\tilde{n}})(\bar{\mathbf{W}}_d^{\text{MMSE}})^H\\
&\quad - (\bar{\mathbf{W}}_d^{\text{MMSE}})^H \tilde{\tilde{\mathbf{H}}}_2 \bar{\mathbf{G}}_r^{\text{BB}} \tilde{\tilde{\mathbf{H}}}_1 \bar{\mathbf{F}}_t^{\text{BB}} - (\tilde{\tilde{\mathbf{H}}}_2 \bar{\mathbf{G}}_r^{\text{BB}} \tilde{\tilde{\mathbf{H}}}_1 \bar{\mathbf{F}}_t^{\text{BB}})^H \bar{\mathbf{W}}_d^{\text{BB}} + \mathbf{I}_{N_s}\\
&= \mathbf{I}_{N_s} - (\tilde{\tilde{\mathbf{H}}}_2 \bar{\mathbf{G}}_r^{\text{BB}} \tilde{\tilde{\mathbf{H}}}_1 \bar{\mathbf{F}}_t^{\text{BB}})^H (\mathbf{A} + \mathbf{R}_{\tilde{n}})^{-1} \tilde{\tilde{\mathbf{H}}}_2 \bar{\mathbf{G}}_r^{\text{BB}} \tilde{\tilde{\mathbf{H}}}_1 \bar{\mathbf{F}}_t^{\text{BB}}.
\end{aligned}
\tag{58}
$$

Based on (58), we can amend our results based on the imperfect channel model, using the same procedure as Section III. For $\bar{\mathbf{G}}_r^{\text{BB}}$, the amended expression is

$$
\begin{aligned}
\bar{\mathbf{G}}_r^{\text{BB}} &= (\mathbf{K}_1 + \lambda^r (\bar{\mathbf{F}}_r^{\text{RF}})^H \bar{\mathbf{F}}_r^{\text{RF}})^{-1} (\tilde{\tilde{\mathbf{H}}}_2)^H (\bar{\mathbf{W}}_d^{\text{MMSE}})^H \bar{\mathbf{V}} (\bar{\mathbf{F}}_t^{\text{BB}})^H\\
&\quad (\mathbf{K}_2 + \sigma_1^2 \bar{\mathbf{W}}_r^{\text{RF}} (\bar{\mathbf{W}}_r^{\text{RF}})^H)^{-1},
\end{aligned}
\tag{59}
$$

where

$$
\bar{\mathbf{V}} = \frac{\mathbb{E}_{\Delta_1,\Delta_2}[\mathbf{E}_{\text{MMSE}}]^{-1}}{\log 2},
$$

21

$$\mathbf{K}_1 := (\bar{\mathbf{W}}_d^{\text{MMSE}} \tilde{\mathbf{H}}_2)^H \bar{\mathbf{V}} \bar{\mathbf{W}}_d^{\text{MMSE}} \tilde{\mathbf{H}}_2 + \text{Tr}(\tilde{\mathbf{\Phi}}_2 (\bar{\mathbf{W}}_d^{\text{MMSE}})^H \bar{\mathbf{V}} \bar{\mathbf{W}}_d^{\text{MMSE}}) \tilde{\mathbf{\Theta}}_2,$$

$$\mathbf{K}_2 := \tilde{\mathbf{H}}_1 \bar{\mathbf{F}}_t^{\text{BB}} (\tilde{\mathbf{H}}_1 \bar{\mathbf{F}}_t^{\text{BB}})^H + \text{Tr}(\bar{\mathbf{F}}_t^{\text{BB}} (\bar{\mathbf{F}}_t^{\text{BB}})^H \tilde{\mathbf{\Theta}}_1) \tilde{\mathbf{\Phi}}_1.$$

For $\bar{\mathbf{F}}_t^{\text{BB}}$, the amended expression is

$$\bar{\mathbf{F}}_t^{\text{BB}} = (\mathbf{T}_1 + \lambda_1^t (\bar{\mathbf{F}}_t^{\text{RF}})^H \bar{\mathbf{F}}_t^{\text{RF}} + \lambda_2^t \mathbf{T}_2)^{-1} (\bar{\mathbf{V}} \bar{\mathbf{W}}_d^{\text{MMSE}} \tilde{\mathbf{H}}_2 \bar{\mathbf{G}}_r^{\text{BB}} \tilde{\tilde{\mathbf{H}}}_1)^H, \tag{60}$$

where

$$\mathbf{T}_1 := \tilde{\mathbf{H}}_1^H (\bar{\mathbf{G}}_r^{\text{BB}})^H \mathbf{B} \bar{\mathbf{G}}_r^{\text{BB}} \tilde{\mathbf{H}}_1 + \text{Tr}(\tilde{\mathbf{\Phi}}_1 (\bar{\mathbf{G}}_r^{\text{BB}})^H \mathbf{B} \bar{\mathbf{G}}_r^{\text{BB}}) \tilde{\mathbf{\Theta}}_1,$$

$$\mathbf{T}_2 := (\mathbf{F}_r^{\text{RF}} \mathbf{G}_r^{\text{BB}} \tilde{\mathbf{H}}_1)^H \mathbf{F}_r^{\text{RF}} \mathbf{G}_r^{\text{BB}} \tilde{\mathbf{H}}_1 + \text{Tr}(\tilde{\mathbf{\Phi}}_1 (\mathbf{F}_r^{\text{RF}} \mathbf{G}_r^{\text{BB}})^H \mathbf{F}_r^{\text{RF}} \mathbf{G}_r^{\text{BB}}) \tilde{\mathbf{\Theta}}_1,$$

$$\mathbf{B} := \tilde{\mathbf{H}}_2^H (\bar{\mathbf{W}}_d^{\text{MMSE}})^H \mathbf{V} \bar{\mathbf{W}}_d^{\text{MMSE}} \tilde{\mathbf{H}}_2 + \text{Tr}(\tilde{\mathbf{\Phi}}_2 (\bar{\mathbf{W}}_d^{\text{MMSE}})^H \mathbf{V} \bar{\mathbf{W}}_d^{\text{MMSE}}) \tilde{\mathbf{\Theta}}_2.$$

Based on the above modifications, our robust baseband design for $\bar{\mathbf{G}}_r^{\text{BB}}$ and $\bar{\mathbf{F}}_t^{\text{BB}}$ is as follows:

1) Calculate the MMSE receiver $\bar{\mathbf{W}}_d^{\text{MMSE}}$ in Eq. (57) and the MMSE matrix $\mathbb{E}_{\Delta_1, \Delta_2}[\mathbf{E}_{\text{MMSE}}]$ in Eq. (58).

2) Update $\bar{\mathbf{V}}$ by setting $\bar{\mathbf{V}} = \frac{\mathbb{E}_{\Delta_1, \Delta_2}[\mathbf{E}_{\text{MMSE}}]^{-1}}{\log 2}$.

3) Fix $\bar{\mathbf{V}}$ and $\bar{\mathbf{F}}_t^{\text{BB}}$, then we find $\bar{\mathbf{G}}_r^{\text{BB}}$ that minimizes $\text{Tr}(\bar{\mathbf{V}} \mathbb{E}_{\Delta_1, \Delta_2}[\mathbf{E}_{\text{MMSE}}])$ under the power constraints. The solution is given by Eq.(59).

4) Fix $\bar{\mathbf{V}}$ and $\bar{\mathbf{G}}_r^{\text{BB}}$, then we find $\bar{\mathbf{F}}_t^{\text{BB}}$ that minimizes $\text{Tr}(\bar{\mathbf{V}} \mathbb{E}_{\Delta_1, \Delta_2}[\mathbf{E}_{\text{MMSE}}])$ under the power constraints. The solution is given by Eq.(60).

After we obtain $\bar{\mathbf{G}}_r^{\text{BB}}$ and $\bar{\mathbf{F}}_t^{\text{BB}}$, we will use MMSE-SIC to design $\bar{\mathbf{W}}_d^{\text{BB}}$, which is the same as what we did in Section III.

V. SIMULATION RESULTS

A. non-robust case

In this section, we consider a relay MIMO system consisting of one source node equipped with a $N_t = 64$ antenna array, a relay node with an $N_r = 32$ antenna array and a destination node with a $N_d = 48$ antenna array unless other number of antennas are specifically mentioned. The number of antennas is chosen from [14] for the purpose of the comparison. For simplicity, we use the channel model in Eq. (9) for channel realization. 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 SNR of the

source-to-relay link and the relay-to-destination are assumed to be the same. In the simulation, we calculate the variances of AWGN noises σ_1 and σ_2 according to the source power and the relay power to maintain the same SNR.
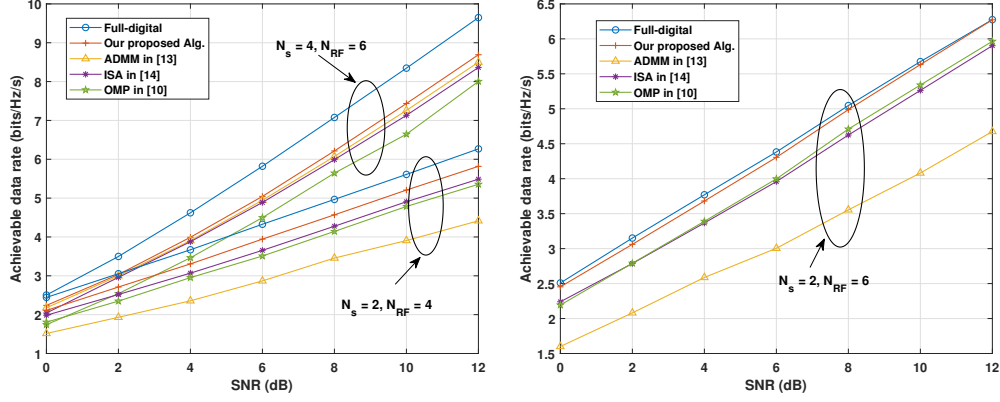

Fig. 4: Achievable rate comparison with $64 \times 32 \times 48$ when $E_s = E_r = N_s$
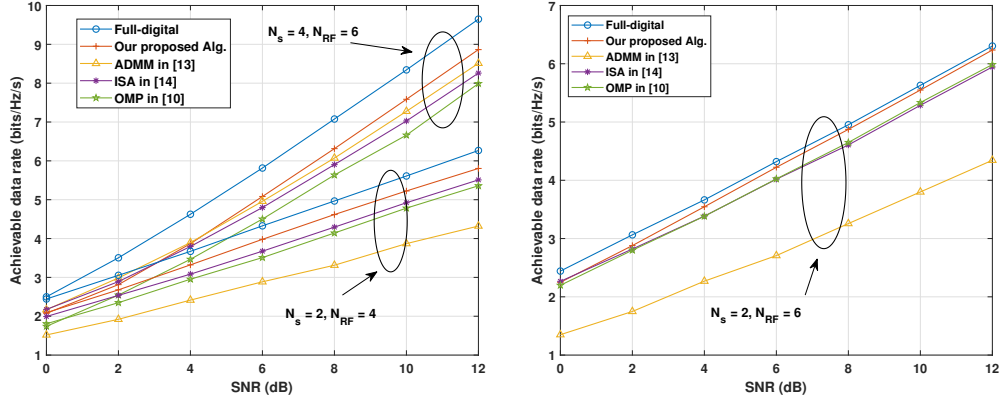

Fig. 5: Achievable rate comparison with $64 \times 32 \times 48$ when $E_s = 2E_r = 2N_s$

In Fig. 4, we equally set the power of source node and the relay node, all to be N_s. We compare our algorithm with the ADMM in [13], the ISA in [14] and the OMP in [10] in terms of the achievable data rate. We use three scenarios: i) the number of data streams is $N_s = 4$ and the number of RF chains is $N_{RF} = 6$; ii) the number of data streams is $N_s = 2$ and the number of RF chains is $N_{RF} = 4$; iii) the number of data streams is $N_s = 2$ and the number of RF chains is $N_{RF} = 6$. The full-digital method is used as a benchmark, where we use the singular matrices of $\mathbf{H}_1$ and $\mathbf{H}_2$ as the precoding/combining matrices. When $N_s = 4$, our algorithm outperforms ADMM by 2%, ISA by 4% and OMP by 9% at SNR $= 12$ dB. When $N_s = 2$ and $N_{RF} = 4$, our

algorithm outperforms ADMM by 32%, ISA by 6% and OMP by 9% at SNR $= 12$ dB. When $N_s = 2$ and $N_{RF} = 6$, our algorithm outperforms ADMM by 34%, ISA by 6 % and OMP by 5% at SNR $= 12$ dB.

In Fig. 5, we set $E_s = 2E_r = 2N_s$. Our proposed algorithm outperforms the other three methods in three scenarios. When $N_s = 4$, our algorithm can provide 4%, 7% and 11% gains over ADMM, ISA and OMP, respectively, at SNR $= 12$ *dB*. When $N_s = 2$ and $N_{RF} = 4$, our algorithm can provide a gain of 34% over ADMM, 5% over ISA and 8% over OMP at SNR $= 12$ dB. When $N_s = 2$ and $N_{RF} = 6$, our algorithm can provide a gain of 44% over ADMM, 5% over ISA and 4% over OMP at SNR $= 12$ dB.
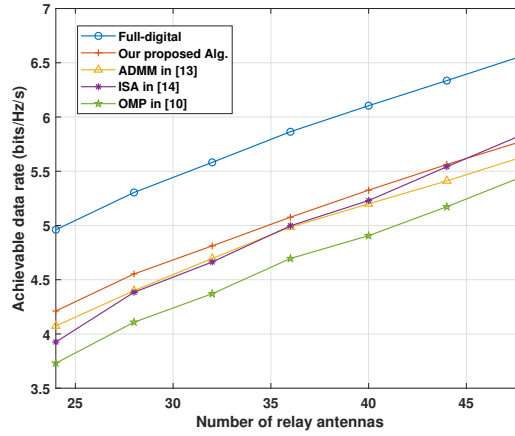
Fig. 6: Achievable rate comparison with different relay antennas when $N_s = 4$, $N_{RF} = 6$ and SNR $= 5$dB

Fig. 6 compares the achievable rate of different algorithms for different number of relay antennas when $N_s = 4$, $N_{RF} = 6$ and SNR $= 5$ dB. The full-digital method is used as a benchmark. As expected, when the number of antennas at the relay node increases, the performance of all different algorithms improves because of the additional antenna gain. Our proposed method has the best achievable rate performance among the four methods except for $N_r = 48$. When $N_r = 48$, ISA has the highest achievable rate among the four methods. However, as the number of antennas at the relay node increases, the complexity of the ISA increases greatly, which will lead to a high power consumption.

Fig. 7 compares the achievable rate for different number of antennas at the destination node when $N_s = 4$, $N_{RF} = 6$ and SNR $= 5$ dB. Similar to Fig. 6, when the number of antennas at the
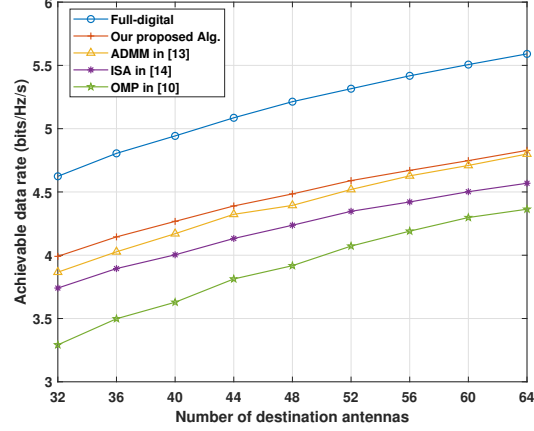
Fig. 7: Achievable rate comparison with different destination antennas when $N_s = 4$, $N_{RF} = 6$ and $SNR = 5$ dB

destination node increases, the performance of all different algorithms improves because of the additional antenna gain. Our proposed method has the best achievable rate performance among the four methods.
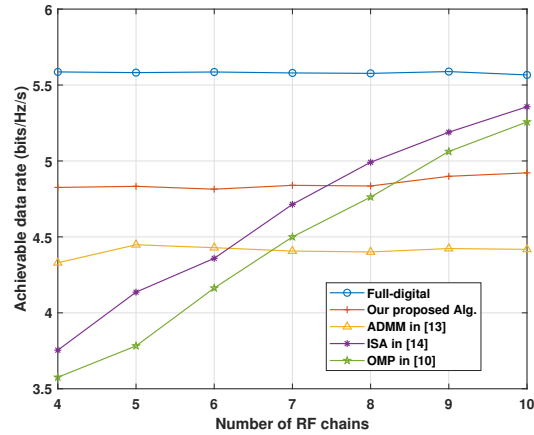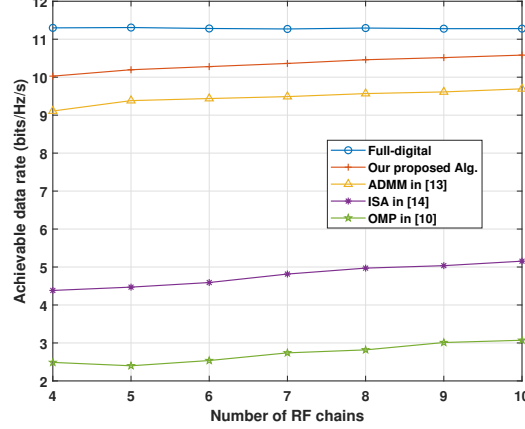

Fig. 8: Achievable rate comparison with different RF chains when $N_s = 4$ and $SNR = 5$ dB using channel model (9)

Fig. 8 compares the achievable rate among the four methods for different number of RF chains when $N_s = 4$ and $SNR = 5$ dB. Since our proposed method is designed to maximize the mutual information between the destination node and the source node after RF precoding/combining,

25

Fig. 9: Achievable rate comparison with different RF chains when $N_s = 4$ and SNR $= 5$ dB using channel model (7)

the gap between our method and the full-digital method is more-or-less fixed, which is caused by the analog processing. However, ISA and OMP are approximation algorithms jointly iterating between the RF and the baseband. Therefore, as the number of RF chains increases, the performance improves. When the number of RF chains is larger than 8, ISA and OMP will outperform our proposed algorithm. However, larger number of RF chains leads to higher complexity and more power consumption. Also, the performance of the approximation algorithms depends on the limited scattering characteristic of the channel. The more sparse the channel is, the better performance the approximation algorithms achieve. In Fig. 8, we use the highly limited scattering channel model in (9), where each scatter only contributes to one path, thus the approximation algorithms have good performance. If we use the general channel model in (7), the performance of approximation algorithms degrades greatly as shown in Fig. 9. In Fig. 9, we set the number of propagation paths N_{cl} in each scatter to be 2 and the number of scatters L to be 20. In this case, the performance of ISA and OMP falls far behind our proposed algorithm.

Fig. 10 shows the convergence performance of different algorithms with respect to the number of iterations. In our algorithm, we update the WMMSE matrix, the digital relay matrix and the digital precoding matrix sequentially in each iteration. In ISA, the digital relay filter, the analog relay receiver and the analog relay precoder are updated sequentially in each iteration. In ADMM, the source node, the relay node and the destination node are optimized alternatively in each iteration. In Fig. 10, our algorithm has the fastest convergence rate while ADMM has the slowest

convergence rate. Moreover, our algorithm has much lower complexity in each iteration compared with ISA. 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 ISA.
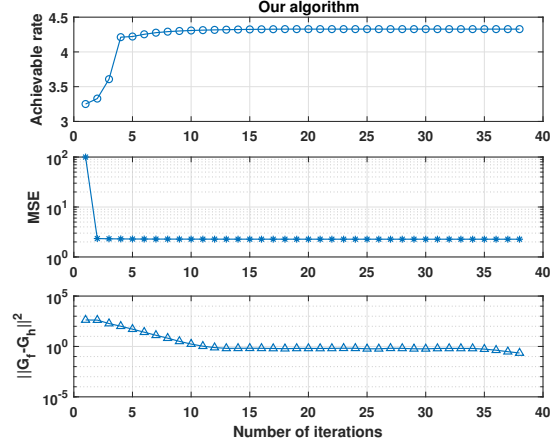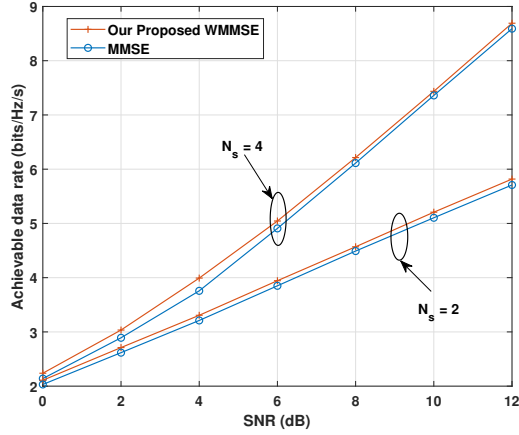

Fig. 10: Convergence rate comparison

Fig. 11: Baseband algorithms comparison

Fig. 11 compares the baseband processing algorithms. Note that we apply the MMSE algorithm only on the baseband, i.e., on the $\tilde{\mathbf{H}}_1$ and $\tilde{\mathbf{H}}_2$. Our proposed WMMSE algorithm outperforms the MMSE algorithm in terms of the achievable data rate since we optimized the mutual information

$I(\mathbf{s}, \mathbf{y_d})$. In fact, if we set our weight matrix to be the identity matrix, our algorithm degenerates to the MMSE algorithm. Therefore, the MMSE algorithm can be considered as a special case of our proposed WMMSE algorithm and our algorithm strictly performs better than MMSE.

B. Robust case

As we described in Section IV, we adopt the channel estimation error model from [18], [20], [24] where the entries of the correlation matrices are selected as $\Phi_1(i,j) = \sigma_{e,1}^2 \beta_1^{|i-j|}$, $\Theta_1(i,j) = \alpha_1^{|i-j|}$, $\Phi_2(i,j) = \sigma_{e,2}^2 \beta_2^{|i-j|}$ and $\Theta_2(i,j) = \alpha_2^{|i-j|}$. Parameters α_1, β_1, α_2 and β_2 are the correlation coefficients and $\sigma_{e,1}^2$ and $\sigma_{e,2}^2$ denote the estimation error covariance. For simplicity, we assume $\alpha_1 = \alpha_2 = \alpha$, $\beta_1 = \beta_2 = \beta$ and $\sigma_{e,1}^2 = \sigma_{e,2}^2 = \sigma_e^2$. The antenna settings are the same as the non-robust part and the number of scatters is set to be 20. The actual channels $\mathbf{H}_1$ and $\mathbf{H}_2$ are generated based on sparse channel model (9) and the estimated channels are generated by $\bar{\mathbf{H}}_1 = \mathbf{H}_1 - \Phi_1^{\frac{1}{2}} \Delta_1 \Theta_1^{\frac{1}{2}}$ and $\bar{\mathbf{H}}_2 = \mathbf{H}_2 - \Phi_2^{\frac{1}{2}} \Delta_2 \Theta_2^{\frac{1}{2}}$.

Fig. 12 shows the effects of the channel estimation error. We provide the performance of our algorithm and those of [10], [14]. For this simulation, we have chosen $\sigma^2 = 0.1$, $\alpha = 0.6$ and $\beta = 0.4$. As shown in Fig. 12, the imperfect channel information will result in severe performance degradation. The achievable data rate of [10], [14] can be decreased to half of what it is for the perfect CSI.

The achievable data rate performances of the proposed robust scheme with various antenna covariance values are depicted in Figs. 13 and 14. When SNR is low, the estimation error can be neglected compared to the noise, therefore the non-robust algorithm achieves good performance which can be even better than that of the robust algorithm. When SNR goes up, the performance of the non-robust algorithm starts to degrade. In Fig. 14, the performance becomes worse than that of the low SNR region for large σ_e^2. Meanwhile, the proposed robust design offers significant gain considering various σ_e^2, which demonstrates the effectiveness of the modified robust transceiver optimization.

In Fig. 15, we compare our robust algorithm with the OMP algorithm in [23]. We set $\alpha = 0$ and $\beta = 0$ for simplicity. The proposed robust design provides a large gain over the algorithm in [23] in all three σ_e^2 settings, showing the advantage of our algorithm.
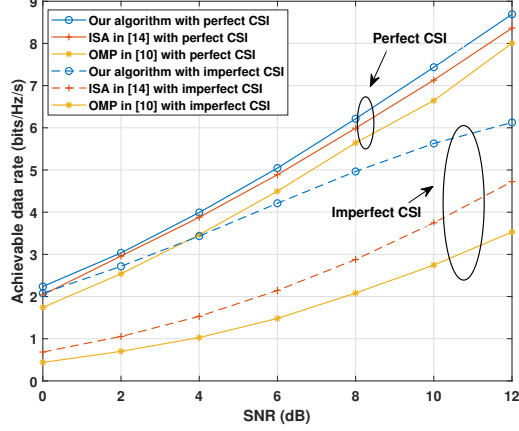
Fig. 12: The effect of imperfect CSI when the error covariance $\sigma_e^2 = 0.1$, $\alpha = 0.6$ and $\beta = 0.4$
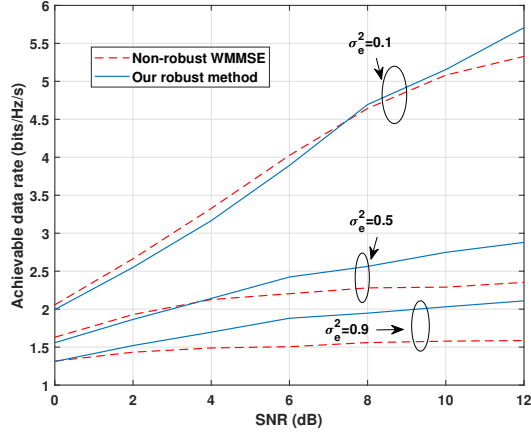

Fig. 13: Achievable rate comparison when $\alpha = 0$, $\beta = 0$

VI. 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 designed the source node and the relay node by making use of the equivalence between maximizing the mutual information and the WMMSE. Given the optimal baseband source and relay filters, we implemented MMSE-SIC for
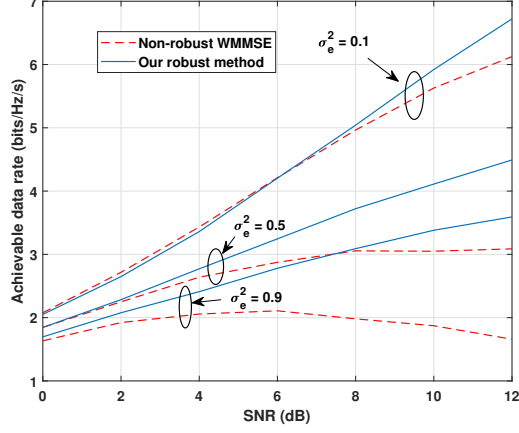
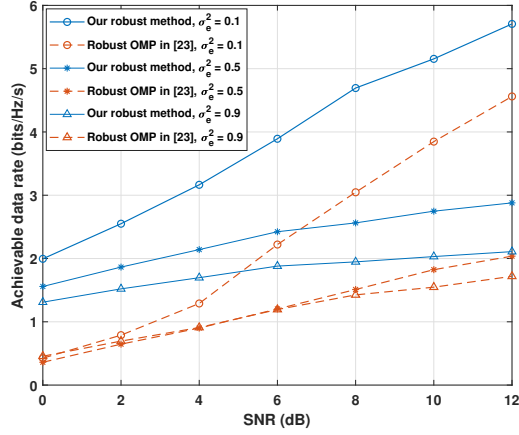Fig. 14: Achievable rate comparison when $\alpha = 0.5$, $\beta = 0.5$

Fig. 15: Achievable rate comparison with the OMP algorithm in [23] when $\alpha = 0$, $\beta = 0$

baseband destination node to obtain the maximal mutual information. In addition, a robust hybrid precoding/combining design was proposed for the imperfect CSI. Simulation results show that our algorithm achieves better performance with lower complexity compared with other algorithms in the literature.

REFERENCES

[1] L. Jiang, X. L. Liu, and H. Jafarkhani, "Hybrid precoding/combining design in mmwave amplify-and-forward MIMO relay networks," in *Proc. of IEEE International Conference on Communications (ICC)*, Shanghai, China, May 2019.

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

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

[4] 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.

[5] 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.

[6] Y. Jing and H. Jafarkhani, "Network beamforming with channel means and covariances at relays," in *Proc. of IEEE International Conference on Communications (ICC)*, Beijing, China, May 2008.

[7] ——, "Network beamforming using relays with perfect channel information," *IEEE Transactions on Information Theory*, vol. 55, no. 6, pp. 2499–2517, Jun. 2009.

[8] 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.

[9] 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.

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

[11] X. Xue, Y. Wang, X. Wang, and T. E. Bogale, "Joint source and relay precoding in multiantenna millimeter-wave systems," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 6, pp. 4924–4937, Oct. 2016.

[12] J.-S. Sheu, "Hybrid digital and analogue beamforming design for millimeter wave relaying systems," *Journal of Communications and Networks*, vol. 19, no. 5, pp. 461–469, Nov. 2017.

[13] C. G. Tsinos, S. Chatzinotas, and B. Ottersten, "Hybrid analog-digital transceiver designs for mmwave amplify-and-forward relaying systems," in *Proc. of 41st IEEE International Conference on Telecommunications and Signal Processing (TSP)*, Athens, Greece, Jul. 2018.

[14] 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.

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

[16] 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.

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

[18] C. Xing, S. Ma, and Y.-C. Wu, "Robust joint design of linear relay precoder and destination equalizer for dual-hop amplify-and-forward MIMO relay systems," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2273–2283, Apr. 2010.

[19] C. Xing, S. Ma, Y.-C. Wu, and T.-S. Ng, "Transceiver design for dual-hop nonregenerative MIMO-OFDM relay systems under channel uncertainties," *IEEE Transactions on Signal Processing*, vol. 58, no. 12, pp. 6325–6339, Dec. 2010.

[20] Y. Rong, "Robust design for linear non-regenerative MIMO relays with imperfect channel state information," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2455–2460, May 2011.

[21] E. Koyuncu and H. Jafarkhani, "Distributed beamforming in wireless multiuser relay-interference networks with quantized feedback," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4538–4576, Jul. 2012.

[22] X. Liu, H. Jafarkhani, and E. Koyuncu, "Amplify-and-forward relay networks with variable-length limited feedback," *IEEE Transactions on Wireless Communications*, vol. 15, no. 11, pp. 7725–7737, Nov. 2016.

[23] Z. Luo, C. Zhan, L. Zhang, and R. Zhang, "Robust hybrid beamforming in millimeter wave relay networks with imperfect CSI," *IEEE Access*, vol. 6, pp. 73 093–73 101, Nov. 2018.

[24] X. Zhang, D. P. Palomar, and B. Ottersten, "Statistically robust design of linear mimo transceivers," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3678–3689, Aug. 2008.

[25] 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.

[26] 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.

[27] P. F. Smulders and L. Correia, "Characterisation of propagation in 60 GHz radio channels," *Electronics & Communication Engineering Journal*, vol. 9, no. 2, pp. 73–80, Apr. 1997.

[28] H. Xu, V. Kukshya, and T. S. Rappaport, "Spatial and temporal characteristics of 60-GHz indoor channels," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 3, pp. 620–630, Apr. 2002.

[29] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Transactions on Signal Processing*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[30] C. A. Balanis, *Antenna theory: analysis and design*. John wiley & sons, 2016.

[31] S. Nadarajah, "A generalized normal distribution," *Journal of Applied Statistics*, vol. 32, no. 7, pp. 685–694, Sep. 2005.

[32] A. K. Gupta and D. K. Nagar, *Matrix variate distributions*. Chapman and Hall/CRC, 2018.