Hardware Impairments Aware Transceiver for Full-Duplex Massive MIMO Relaying

Xiaochen Xia, Dongmei Zhang, Kui Xu, Member, IEEE, Wenfeng Ma, and Youyun Xu, Senior Member, IEEE

Abstract—This paper studies the massive MIMO full-duplex relaying (MM-FDR), where multiple source-destination pairs communicate simultaneously with the help of a common full-duplex relay equipped with very large antenna arrays. Different from the traditional MM-FDR protocol, a general model where sources/destinations are allowed to equip with multiple antennas is considered. In contrast to the conventional MIMO system, massive MIMO must be built with low-cost components which are prone to hardware impairments. In this paper, the effect of hardware impairments is taken into consideration, and is modeled using transmit/receive distortion noises. We propose a low complexity hardware impairments aware transceiver scheme (named as HIA scheme) to mitigate the distortion noises by exploiting the statistical knowledge of channels and antenna arrays at sources and destinations. A joint degree of freedom and power optimization algorithm is presented to further optimize the spectral efficiency of HIA based MM-FDR. The results show that the HIA scheme can mitigate the “ceiling effect” appears in traditional MM-FDR protocol, if the numbers of antennas at sources and destinations can scale with that at the relay.

Index Terms—Massive MIMO full-duplex relaying, hardware impairments, transceiver design, joint degree of freedom and power optimization, achievable rate.

I. INTRODUCTION

In multi-user MIMO systems, one main challenge is the increased complexity and energy consumption of the signal processing to mitigate the interference between multiple co-channel users. To achieve energy efficient transmission, the multi-user MIMO system with very large antenna arrays at each base station (known as “massive MIMO” system) has been advocated recently [1]. The key result is that, with very large antenna arrays at each base station, both the intracell and intercell interferences can be substantially reduced with simple linear beamforming (BF) processing [1], [2].

On the other hand, full-duplex relaying (FDR) is a promising approach to improve the spectral efficiency (SE) of relaying network while retains the merits of half-duplex relaying (HDR) (e.g., path loss reduction). In FDR, the relay transmits and receives simultaneously at the same frequency and time, but at the cost of a strong echo interference (EI) due to signal leakage between the relay output and input [3]. To mitigate EI, three approaches have been investigated, i.e., 1) passive cancellation, 2) time-domain cancellation [3] and 3) spatial suppression [5], [6]. The passive cancellation relies on a combination of path loss, cross-polarization and antenna directionality [4]. The time-domain cancellation relies on the fact that EI signal is known at full-duplex node. Thus, it can be regenerated and removed in time-domain [3]. In spatial suppression, EI is mitigated with the multiple transmit/receive antennas by approaches such as null-space projection [5]. Inspired by these works, a number of works have dedicated to the study of FDR protocol on both theory and testbed (See [7], [8] and the references therein). To achieve spectral and energy efficient transmissions of multiple source-destination pairs, recent works sought to incorporate both HDR [9], [10] and FDR [11]-[13] with massive MIMO.

However, the aforementioned works on massive MIMO are actually based on the assumption that the base stations or relays are equipped with a large number of high-quality transmit/receive radio frequency (RF) chains (which are expensive and power-hungry). In contrast to conventional MIMO system (e.g., at most 8 antennas in LTE system), massive MIMO must be built with low-cost components [14] since the deploy cost and energy consumption of circuits will increase dramatically as the number of antennas grows very large. Such low-cost components are prone to hardware imperfections (e.g., phase noise, nonlinear power amplifier, I/Q imbalance, nonlinear low-noise amplifier and ADC impairments), which must be considered in the design of practical massive MIMO system.

This paper focuses on the transceiver design for massive MIMO full-duplex relaying (MM-FDR) with hardware impairments. The effect of hardware impairments is modeled using transmit/receive distortion noises [15], [16]. There are several challenges in the design of practical transceiver scheme in the considered system. The first is: how to deal with the EI cancellation without instantaneous EI channel? EI cancellation is a critical problem in MM-FDR transceiver design which is not only important to reduce EI power, but also useful to reduce distortion noises caused by hardware impairments at the relay (as will be shown in section IV). Different from FDR with small-scale relay antenna arrays [8]-[8], the instantaneous EI channel is usually not easy to obtain in MM-FDR. This is because the learning of EI channel requires training sequence with length not less than the number of relay antennas [2], which is prohibitive in MM-FDR as the channel coherent time is limited. The lack of EI channel makes traditional EI cancellation techniques (e.g., time-domain cancellation and spatial suppression) difficult to apply. Although passive cancellation does not relies on EI channel...
estimation, it usually cannot provide satisfactory performance when used alone [8]. Another problem is: How to suppress distortion noises caused by hardware impairments at sources and destinations. Different from multi-user interference (MUI), the distortion noise caused by transmit imperfection of source can be viewed as an interference signal with the same channel as transmit data. Thus, it cannot be suppressed by relay antenna arrays during coherent combining (Similar problem appears in reception at destinations). This causes performance ceiling on achievable rate as the number of relay antennas grows large, which degrades the gain of massive MIMO significantly.

In this paper, we propose practical transceiver scheme for MM-FDR with hardware impairments considering the above problems. Different from the traditional MM-M-FDR protocol [12] (where sources and destinations are equipped with single antenna), we consider a general model where sources and destinations are allowed to equip with multiple antennas. The contributions are summarized as follows:

- We first examine the limitation of traditional MM-M-FDR protocol under hardware impairments. In particular, we derive the upper bound on end-to-end achievable rate for traditional MM-M-FDR protocol with linear processing at the relay. The bound reveals that the achievable rate is limited by the hardware impairments at the sources and destinations, and performance ceiling appears as the number of relay antenna grows large. The result also implies that it is impossible to cancel the “ceiling effect” with linear processing, if sources and destinations are only equipped with single antenna.

- Based on the upper bound analysis, we propose a hardware impairments aware transceiver scheme (HIA scheme) to mitigate the distortion noises by exploiting the statistical channel knowledge and antennas arrays of sources/destinations. The scheme needs no instantaneous knowledge of El channel. The asymptotic end-to-end achievable rate of MM-FDR with HIA scheme (HIA-M-M-FDR) is derived and the scaling behaviors of MUI, El and distortion noises are determined.

- A joint degree of freedom and power optimization (JDPO) algorithm is presented to further improve the SE of HIA-M-MM-FDR.

The paper is organized as follows. In section II, we review the related work. The system model is described in section III. The upper bound of achievable rate is analyzed in section IV. The HIA scheme is proposed in section V and the JDPO algorithm is presented in section VI. Simulation results are presented in section VII and some conclusions will be drawn at last. Notations: \( E(\cdot) \) and \( \text{var}(\cdot) \) denote the expectation and variance. \( I_n \) is \( n \times n \) identity matrix. \( (\cdot)^* \), \( (\cdot)^T \) and \( (\cdot)^H \) denote conjugate transpose, transpose and conjugate-transpose, respectively. \( \rho(A) \), \( \text{Tr}(A) \), \( |A|_{ij} \), \( \lambda_l(A) \) and \( u_l(A) \) denote the spectral radius, trace, \( (i, j) \)th element of matrix, \( l \)th largest eigenvalue and eigenvector with respect to \( l \)th largest eigenvalue of matrix \( A \), respectively. \( a \) scales with \( b \) means \( 0 < \lim_{b \to \infty} \frac{a}{b} < \infty \).

II. RELATED WORK

The design and performance of using unlimited number of antennas at the base station in cellular system were first considered for independent antennas [1], [2], and soon extended to the scenarios with spatial correlated antennas [17]-[19]. The asymptotic SINR for single cell cellular system was analyzed in [2]. It has been shown that, with very large antenna arrays at the base station, a deterministic SINR (also called the “deterministic equivalent” of SINR) which depends only on the large-scale fading of channels can be achieved, if the transmit power is scaled by \( N \) with perfect CSI and \( \sqrt{N} \) with imperfect CSI (\( N \) denotes the number of base station antennas). The authors in [17]-[20] have done considerable work to derive the “deterministic equivalent” of SINR for massive MIMO system with spatial correlated antennas.

In the field of massive MIMO relaying, the SE and EE of HDR with very large relay antenna arrays were investigated in [10], [9]. The MM-FDR with decode-and-forward (DF) relay was first introduced in [11], [12], and analyzed in end-to-end achievable rate as linear processing is employed. The asymptotic performance of amplify-and-forward based MM-FDR was considered in [13] and the scaling behavior of the infinitely repeating echo interference was determined.

Only a few works considered the effect of hardware impairments on massive MIMO system. A constant envelope signal design has been proposed in [21] to facilitate the use of power-efficient RF power amplifiers. The authors in [22] presented a low peak-to-average-power ratio (PAR) precoding solution to enable efficient implementation using non-linear RF components in massive MIMO system. In FDR with small-scale antennas, the optimal precoding under limited ADC dynamic range was studied in [15], [16]. However, these works cannot provide much insight for the effect of hardware impairments on MM-FDR.

III. SYSTEM MODEL

Consider the network with \( K \) source-destination (S-D) pairs and a single full-duplex relay \( R \), where source \( S_k \) wishes to communicate with destination \( D_{k} (k \in \{1, \cdots, K\}) \) with the help of \( R \). The relay adopts the DF policy. It is assumed that the sources and destinations are equipped with \( N_S \) and \( N_D \) antennas respectively, while the relay is equipped with \( N_R + N_T \) antennas (\( N_R \) for reception and \( N_T \) for transmission). We are interested in the large-\( (N_R, N_T) \) regime, i.e., \( \min\{N_R, N_T\} \to \infty \). \( N_S, N_D \) and \( K \) can be either fixed or scale with \( \min\{N_R, N_T\} \).

Let \( H_{SR,k} \in \mathbb{C}^{N_R \times N_S} \) be the channel matrix from \( S_k \) to receive antenna array of relay and let \( H_{RD,k} \in \mathbb{C}^{N_T \times N_D} \) be the channel matrix from \( D_k \) to transmit array of relay. Let \( H_{EI} \in \mathbb{C}^{N_R \times N_T} \) denote the EI channel matrix between transmit and receive arrays of relay. The spatial correlation of each MIMO channel is characterized by the Kronecker model [23]. Thus, \( H_{SR,k} \) can be expressed as \( H_{SR,k} = \sqrt{\beta_{SR,k}} C_{SR,k} X_{SR,k} G_{SR,k}^{1/2} \), where \( \beta_{SR,k} \) represents the large-scale fading, \( C_{SR,k} \in \mathbb{C}^{N_R \times N_R} \) and \( G_{SR,k} \in \mathbb{C}^{N_S \times N_S} \) characterize the spatial correlation of received signals across receive array of relay and that of transmitted signals across transmit array of \( S_k \). \( X_{SR,k} \in \mathbb{C}^{N_R \times N_D} \) consists of the random components of channels whose elements are i.i.d with distribution \( \mathcal{CN}(0,1) \). Based on the
same model, the channel matrices $H_{RD,k}$ and $H_{EI}$ can be expressed as $H_{RD,k} = \sqrt{\beta_{RD,k}} C_{RD,k}^{1/2} x_{RD,k} C_{RD,k}^{1/2}$ and $H_{EI} = \sqrt{\beta_{EI}} C_{EI}^{1/2} x_{EI} C_{EI}^{1/2}$. To facilitate the analysis, we assume the following conditions on correlation matrices, i.e., $\forall C \in \{C_{SR,k}, C_{SR,k}, C_{RD,k}, C_{RD,k}, C_{EI}, C_{EI}\}$, where $k \in \{1, \ldots, K\}$,

- **A1**: The spectral radius of $C$ is bounded by a constant, i.e., $\rho(C) \leq C$.
- **A2**: $C$ is a Hermitian and Toeplitz matrix and has unit diagonal elements.

The former is a common assumption in the studies of massive MIMO which follows from energy conservation [17] and A2 corresponds to the case of uniform linear array (ULA) [23].

To characterize the effect of hardware imperfections, we adopt the new signal model from [15], [16].

1) Imperfect Transmit RF Chain: We model the effect of imperfect transmit RF chain by adding, per transmit antenna, an independent zero-mean Gaussian “distortion noise”, whose variance is proportional to the signal power transmitted at that antenna. The experimental results in [24], [25] have shown that the independent Gaussian distortion noise model closely captures the joint effect of imperfect components in transmit RF chain. Let $x_{S,k}[u] \in \mathbb{C}^{N_s \times 1}$ and $x_{R}[u] \in \mathbb{C}^{N_r \times 1}$ denote the transmit vectors of source $S_k$ and relay at time instant $u$. Based on the above model, the distorted transmit signals can be expressed as

$$x_{S,k}[u] = x_{S,k}[u] + t_{S,k}[u]$$

and

$$x_{R}[u] = x_{R}[u] + t_{R}[u]$$

where the distortion noises $t_{S,k}[u] \sim \mathcal{CN}(0, \nu_{S,k} \text{diag}(E[x_{S,k}[u] x_{S,k}[u]^H]))$ and $t_{R}[u] \sim \mathcal{CN}(0, \nu_{R} \text{diag}(E[x_{R}[u] x_{R}[u]^H]))$. Note that $\nu_{S,k} > 0$ ($\nu_{R} > 0$) characterizes the level of transmit distortion. For example, $\nu_{S,k} = 0$ ($\nu_{R} = 0$) corresponds to the conventional assumption of perfect transmit RF chains. The quality of transmit RF chains degrades as $\nu_{S,k}$ ($\nu_{R}$) increases.

2) Imperfect Receive RF Chain: We model the effect of imperfect receive RF chain by adding, per receive antenna, an independent zero-mean Gaussian “distortion noise”, whose variance is proportional to the signal power received at that antenna. More precisely, let $y_{R}[u] \in \mathbb{C}^{N_r \times 1}$ and $y_{D,k}[u] \in \mathbb{C}^{N_o \times 1}$ be the undistorted received signals of the relay and destination $D_k$ at time instant $u$, the distorted received signals can be expressed as

$$y_{R}[u] = y_{R}[u] + r_{R}[u]$$

and

$$y_{D,k}[u] = y_{D,k}[u] + r_{D,k}[u]$$

where the distortion noises $r_{R}[u] \sim \mathcal{CN}(0, \mu_{R} \text{diag}(E[y_{R}[u] y_{R}^H[u]]))$ and $r_{D,k}[u] \sim \mathcal{CN}(0, \mu_{D,k} \text{diag}(E[y_{D,k}[u] y_{D,k}^H[u]]))$. $\mu_{R} > 0$ ($\mu_{D,k} > 0$) characterizes the level of receive imperfection. The experimental studies in [26] have shown that the independent Gaussian noise model is a good approximation to the joint effect of imperfect components in receive RF chain.

At time instant $u$, all sources transmit signals $x_{S,k}[u]$ ($k = 1, \ldots, K$) to the relay simultaneously. Meanwhile, the relay broadcasts the decoded signals to the destinations. Based on the models in (1) and (2), the received signals at the relay and $D_k$ can be expressed as

$$y_{R}[u] = \sum_{k=1}^{K} H_{SR,k} (x_{S,k}[u] + t_{S,k}[u])$$

$$+ H_{EI} (x_{R}[u] + t_{R}[u]) + r_{R}[u] + n_{R}[u]$$

$$y_{D,k}[u] = H_{RD,k}^H (x_{R}[u] + t_{R}[u]) + r_{D,k}[u] + n_{D,k}[u]$$

(3)

To keep the complexity low, it is assumed that a single data stream is transmitted at each source and each node employs only linear processing. In particular, $S_k$ transmits the unit-power symbol $s_k[u]$ using the unitary BF vector $p_{S,k}$. Therefore, the transmit vector of $S_k$ can be expressed as $x_{S,k}[u] = \sqrt{E_{S,k}} p_{S,k} s_k[u]$, where $E_{S,k}$ denotes the transmit power. Relay combines the received signal by multiplying the receive BF matrix $W_{R} = [w_{R,1}, \cdots, w_{R,K}]$, i.e., $y_{R}[u] = W_{R} y_{R}[u]$. The $k$th element of $y_{R}[u]$

$$y_{R,k}[u] = \sqrt{E_{S,k}} w_{R,k}^H H_{SR,k} p_{S,k} s_k[u]$$

$$+ \sqrt{E_{S,j}} \sum_{j=1,j \neq k}^{K} w_{R,j}^H H_{SR,j} p_{S,j} s_j[u] + w_{R,k}^H \sum_{j=1}^{K} H_{SR,j} t_{S,j}[u]$$

$$+ w_{R,k}^H H_{EI} (x_{R}[u] + t_{R}[u]) + w_{R,k}^H r_{R}[u] + w_{R,k}^H n_{R}[u]$$

(4)

is used to decode the symbol of $S_k$. The relay forwards the decoded symbol using transmit BF matrix $W_{T} = [w_{T,1}, \cdots, w_{T,K}]$. The transmit vector of the relay can be expressed as $x_{R}[u] = W_{T} A_{R}^{1/2} s_v[u - d]$, where $s_v[u - d] = [s_1[u - d], \cdots, s_K[u - d]]^T$ and $A_{R} = \text{diag}(E_{R,1}, \cdots, E_{R,K})$ is the power allocation matrix at the relay. $d$ denotes the processing delay of the relay. To meet the relay’s power constraint, $x_{R}[u]$ must satisfy

$$\text{Tr}(E[y_{R}[u] x_{R}^H[u]]) = \sum_{i=1}^{K} E_{R,i} \leq E_{R,\text{max}}.$$ $D_k$ uses the unitary BF vector $p_{D,k}$ to combine the received signal $y_{D,k}[u]$. The combined signal is expressed as

$$y_{D,k}[u] = \sqrt{E_{R,k}} p_{D,k}^H H_{RD,k}^H w_{T,k} s_v[u - d]$$

$$+ p_{D,k}^H \sum_{j=1,j \neq k}^{K} \sqrt{E_{R,j}} H_{RD,j}^H w_{T,j} s_j[u - d]$$

$$+ p_{D,k}^H H_{RD,k}^H t_{R}[u] + p_{D,k}^H r_{D,k} + p_{D,k}^H n_{D,k}[u]$$

(5)

IV. LIMITATION OF MM-FDR WITH SINGLE ANTENNA SOURCES AND DESTINATIONS

This section analyzes the upper bound on achievable rate of MM-FDR with single2 antenna sources and destinations. The goal is to reveal the fundamental limitation of previous MM-FDR protocol with single antenna sources and destinations in combating distortion noises. Moreover, the bound provides a tool for understanding the impact of practical transceiver scheme to mitigate the distortion noises, when multiple antennas are
A. Upper Bound on Achievable Rate

As $N_S = N_D = 1$, we set $p_{S,k} = p_{D,k} = 1$, replace $H_{i,k}$ ($i \in \{SR, RD\}$) with $H_{i,k}$ and let $H_i = [h_{i,1}, \ldots, h_{i,K}]$ and $t_S[u] = [t_{S,1}^T[u], \ldots, t_{S,K}^T[u]]^T$. We consider a block-fading channel with coherence time $T$ (symbol times), where $\tau$ are used for uplink training of each source/destination, and the remaining $T - 2K\tau$ are used for data transmission. The upper bound on achievable rate is obtained by assuming perfect knowledge of $h_{SR,k}, h_{RD,k}$ and $H_{EI}$ can be provided with pilot signals, and meanwhile, the MUs (i.e., the second terms of right-hand side of (4) and (5)) and EI term $w_R^H H_{EI} x_R[u]$ can be somehow cancelled. This gives us the following upper bound on end-to-end achievable rate

$$R_{\text{Upper}} = \frac{T}{T} \log_2 \left( 1 + \frac{E_{S,k} h_{SR,k}^H h_{SR,k} w_{R,k}^T h_{R,k}^H Q w_{R,k}}{h_{R,k}^H \Theta R h_{R,k} + \mathbb{E} \left[ \|r_{D,k}[u]\|^2 \right] + 1} \right)$$

(7)

where $Q$ is given as $Q = H_{SR} \Theta R h_{SR,k}^H h_{SR,k} + H_{EI} \Theta R h_{EI} + \Theta R + I_{N_R}$ with $\Theta R = \mathbb{E} [t_S[u] t_S^T[u]]$ and $\Theta R = \mathbb{E} [t_R[u] t_R^T[u]]$ denoting the covariance matrices of distortion noises $t_S[u], t_R[u]$ and $r_R[u]$, respectively.

1) Achievable Rate of $S_k \rightarrow R$ Channel: Since $w_{R,k}$ is only related with $R_{SR,k}$, it can be optimized separately to maximize $R_{SR,k}$. From (7), the optimization problem is equivalent to the generalized Rayleigh quotient problem, which can be solved as $w_{R,k} = Q^{-1} h_{SR,k}$. The resultant upper bound on $R_{SR,k}$ (denoted by $R_{\text{Upper}}^{\text{Upper}}$) can be expressed as

$$R_{\text{Upper}}^{\text{Upper}} = \frac{T}{T} \log_2 \left( 1 + \frac{E_{S,k} h_{SR,k} Q^{-1} h_{SR,k} w_{R,k}^T h_{R,k}^H Q w_{R,k}}{h_{R,k}^H \Theta R h_{R,k} + \mathbb{E} \left[ \|r_{D,k}[u]\|^2 \right] + 1} \right)$$

(8)

where $Q_k$ is defined as $Q_k = Q - \nu_S k p_{S,k} h_{SR,k} h_{SR,k}^H$ and is independent of $h_{SR,k}$. The second step follows from the matrix inversion lemma [27].

Theorem 1: Assuming that the assumptions A1 and A2 hold. As $N_S = N_D = 1$, the achievable rate of $S_k \rightarrow R$ channel in the large-$(N_R, N_T)$ regime is bounded as

$$R_{SR,k} \leq R_{\text{Upper}} = \frac{T}{T} \log_2 \left( 1 + \frac{E_{S,k} \beta_{SR,k} \Theta_{SR,k} (C_{SR,k} \Psi_k)}{1 + \nu_S k E_{S,k} \beta_{SR,k} \Theta_{SR,k} (C_{SR,k} \Psi_k)} \right)$$

(9)

where $\rho = \beta_{EI} \Theta_{EI} \Psi_k + 1$, $\Psi_k(\rho)$ is determined by the following fixed-point algorithm with $\rho_k(\rho) = \frac{1}{\rho}$

$$\Psi_k(\rho) = \left( \frac{1}{N_R} \sum_{l=1,l \neq k}^K \frac{\nu_S \beta_{SR,l} C_{SR,l}}{1 + \delta_{s,l,k}(\rho)} + \Theta_R + \rho I_{N_R} \right)^{-1}$$

$$\Theta_R = \mu_R \sum_{l=1}^K \nu_S \beta_{SR,l} I_{N_R} + \beta_{EI} \Theta_{EI} \Psi_k + \mu_R I_{N_R}$$

$$\delta_{s,l,k}(\rho) = \lim_{n \rightarrow \infty} \delta_{s,l,k}^{(n)}(\rho)$$

$$\delta_{s,l,k}^{(n)}(\rho) = \frac{\nu_S \beta_{SR,l} \Theta_{SR,l}}{N_R} \Theta_{EI} \left( \frac{1}{N_R} \sum_{l=1,l \neq k}^K \frac{\nu_S \beta_{SR,l} C_{SR,l}}{1 + \delta_{s,l,k}^{(n-1)}(\rho)} + \Theta_R + \rho I_{N_R} \right)^{-1}$$

Proof: See Appendix-B.

The convergence of the fixed-point algorithm in Theorem 1 has been proved in [28].

2) Achievable Rate of $R \rightarrow D_k$ Channel: From (7), the achievable rates of $S_k \rightarrow R$ and $R \rightarrow D_k$ channels are coupled through $\Theta_R = \mathbb{E} [t_R[u] t_R^T[u]] = \mathbb{E} \left[ W_T \Lambda_R W_T^H \right]$, which makes the design of $w_{R,k}$ very challenging. However, with the following theorem we show that this coupling disappears in the large-$(N_R, N_T)$ regime, which allows a simple upper bound for achievable rate of $R \rightarrow D_k$ channel.

Theorem 2: Assuming that the assumptions A1 and A2 hold. As $N_S = N_D = 1$, the achievable rate of $S_k \rightarrow R$ channel is independent of $w_{R,k}$. The achievable rate of $R \rightarrow D_k$ channel in the large-$(N_R, N_T)$ regime is bounded as (11), which is achieved by the eigen BF $w_{R,k} = h_{RD,k}/\|h_{RD,k}\|$.

Proof: See Appendix-C.

With Theorems 1 and 2 the upper bound of end-to-end achievable in the large-$(N_R, N_T)$ regime can be expressed as

$$R_{\text{Upper}}^{\text{Upper}} = \min \left( R_{\text{Upper}}^{\text{Upper}}^{SR,k}, R_{\text{Upper}}^{\text{Upper}}^{RD,k} \right)$$

(12)

Remark 1: The bound in (12) is derived by assuming linear processing and fixed transmit powers at sources and relay, which is in general not capacity achieving. One exception is when $K \ll \min\{N_R, N_T\}$ and $\min\{N_R, N_T\} \rightarrow \infty$. From Theorem 2 the design of receive and transmit BF matrices is decoupled as $\min\{N_R, N_T\} \rightarrow \infty$. Moreover, the distortion noises are assumed to be circularly symmetric complex Gaussian distributed and independent of the desired signal. Thus, based on the results in [29, 30], linear processing along with power control is sufficient to achieve the capacity. When $K \ll \min\{N_R, N_T\}$, the terms $\frac{1}{N_R} \sum_{l=1,l \neq k}^K \frac{\nu_S \beta_{SR,l} C_{SR,l}}{1 + \delta_{s,l,k}(\rho)}$ and
By examining (13) and the expression of \( \Theta \) in (9), the bounds in (12) are expressed as

\[
\frac{1}{N_R} \sum_{j=1, j \neq k}^{K} \frac{\nu_{S,j} E_{S,j} \beta_{S,R,j} C_{SR,j}}{1 + \delta_{k,l}^{(n)}(\rho)} \]  

in (10) \( \delta_{k,l}^{(n)}(\rho) \) and (12) vanish. By neglecting the low order terms (in \( \min\{N_R, N_T\} \)), the bound in (12) reduces to \( \frac{2 - 2K_T}{T} \log_2 \left( 1 + \min\{1 \} \right) \), which is independent of \( E_{S,k} \) and \( E_{R,k} \). Thus, (12) is the upper bound for capacity of \( S_k \rightarrow R \rightarrow D_k \) channel as \( K \ll \min\{N_R, N_T\} \) and \( \min\{N_R, N_T\} \rightarrow \infty \).

### B. Limitation with Hardware Impairments

Different from Remark 1, we consider the general case in which \( K \) can be either fixed or scales with \( \min\{N_R, N_T\} \). With Theorem 1 and Theorem 2, the effect of hardware impairments is still hard to analysis since \( \Psi_k \) is not in closed-form. To make the analysis tractable, we assume \( C_{SR,l} = C_{SR} \forall l \in \{1, \cdots, K\} \). Using eigenvalue decomposition

\[
C_{SR} = U \Sigma_{SR} U^H \quad \text{on } \delta_{k,l}^{(n)}(\rho) \quad \text{in (10)},
\]

we have

\[
\delta_{k,l}^{(n)}(\rho) = \frac{\nu_{S,l} \rho_{S,l} \beta_{S,R,l}}{N_R} \text{Tr} (\Sigma_{SR})
\]

\[
\times \left( \frac{1}{N_R} \sum_{j=1, j \neq k}^{K} \frac{\nu_{S,j} \rho_{S,j} \beta_{S,R,j}}{1 + \delta_{k,l}^{(n-1)}(\rho)} \Sigma_{SR} + \Theta^{R}_{\rho} + \rho I_{N_R} \right)^{-1}
\]

By examining (13) and the expression of \( \Theta^{R}_{\rho} \) in (10), \( \delta_{k,l}^{(n)}(\rho) \) is an \( O(1) \) term. Note that this is valid for arbitrary \( n > 0 \). Thus, we can conclude that \( \delta_{k,l}(\rho) = O(1) \). Replacing \( \delta_{k,l}(\rho) \) in \( \Psi_k \), the symbol \( O(1) \) and substituting the result into (9), \( R_{SR,k} \) can be written as (14).

For convenience, we further assume

\[
\{\beta_{S,R,l}, \beta_{R,D,l}, E_{S,l}, E_{R,l}, \nu_{S,l}, \nu_{R,l}\} = \{\beta_{S,R}, \beta_{R,D}, E_{S}, E_{R}, \nu_{S}, \nu_{R}\}.
\]

Note that this assumption does not affect the basis conclusions of the analysis. By neglecting the low order terms (in \( \min\{N_R, N_T\} \)) in (14) and (11), the upper bound on end-to-end achievable rate reduces to

\[
R_{k}^{\text{Upper}} = \frac{T - 2K_T}{T} \log_2 \left( 1 + \min \left\{ \frac{1}{\nu_{S} + \frac{K}{N_R} \nu_{R} \rho_{S,1} E_{S}/E_{S,R}} \right\} \right)
\]

One can make several observations from (15):

1) As \( K \ll \min\{N_R, N_T\} \), the effect of hardware impairments at the relay and EI disappears in the large-(\( N_R, N_T \)) regime, and the upper bound converges to that in Remark 1.

The end-to-end achievable rate is limited by distortion noises caused by hardware impairments at sources and destinations. The explanation is that the transmit distortion can be viewed as an interference signal the same channel as data of \( S_k \) from

\[
R_{RD,k} \leq R_{k}^{\text{Upper}} = \frac{T - 2K_T}{T} \log_2 \left( 1 + \min \left\{ \frac{1}{\nu_{S} + \frac{K}{N_R} \nu_{R} \rho_{S,1} E_{S}/E_{S,R}} \right\} \right)
\]

Thus, it cannot be suppressed after combining by \( w_{RD,k} \). The power of distortion due to receive imperfection of \( D_k \) is proportional to that of desired receive signal (after eigen BF) and also cannot be reduced by the transmit BF scheme of relay. Thus, there is a finite ceiling on end-to-end achievable rate as \( \min\{N_R, N_T\} \rightarrow \infty \). In fact, this poses a major limitation on MM-FDR with single antenna sources and destinations, which degrades the gain of massive array of relay greatly.

2) As \( K \) scales with \( \min\{N_R, N_T\} \), the effect of hardware impairments at the relay and EI is not negligible. Since the EI channel is typically stronger than desired channels, the \( S_k \rightarrow R \) channel becomes the bottleneck of end-to-end achievable rate. This indicates that it is of great importance to suppress EI to avoid the bottleneck effect.
only on the statistical knowledge of EI channels. With \cite{[4]}, we design an average EI power minimization problem, i.e.,

\[
\left( \hat{P}_R, \hat{P}_T \right) = \arg \min_{P_R, P_T} \mathbb{E} \left[ \left\| W_{R,k} \right\|^2 \left( W_{T}^T \Lambda_R W_{T}^H \right) \right.
\]

\[
\left. \left( \Tr(P_R^H H_{EI} P_T P_T^H H_{EI}^H P_R) \right) + \nu_R \sum_{l=1}^{K} E_{R,k} \left\| W_{R,l} \right\|^2 \left( \Tr(P_R^H H_{EI} H_{EI}^H P_R) \right) \right].
\]

Then the problem \cite{[16]} can be rewritten approximately as

\[
\left( \hat{P}_R, \hat{P}_T \right) = \arg \min_{P_R, P_T} \left\{ \Tr(W_{T}^T \Lambda_R W_{T}^H) \right. \times \mathbb{E} \left[ \Tr(P_R^H H_{EI} P_T P_T^H H_{EI}^H P_R) \right]
\]

\[
\left. + \nu_R \sum_{l=1}^{K} E_{R,k} \mathbb{E} \left[ \Tr(P_R^H H_{EI} H_{EI}^H P_R) \right] \right\}
\]

\[
= \arg \min_{P_R, P_T} \left\{ \Tr(W_{T}^T \Lambda_R W_{T}^H) \times \left( \Tr(P_R^H \hat{C}_{EI} P_T) \right)
\]

\[
\left. + \nu_R \sum_{l=1}^{K} E_{R,k} \Tr(P_R^H \hat{C}_{EI} P_R) + \nu_R N_T \sum_{l=1}^{K} E_{R,k} \Tr(P_R^H C_{EI} P_R) \right\}
\]

(18)

The multiplicative term \( \left\| W_{R,k} \right\|^2 \) has been removed since it has no effect on the optimal solution. The second step follows from a similar derivation of \cite{[55]}. From \cite{[18]}, the columns of \( \hat{P}_R \) (\( \hat{P}_T \)) are composed of the eigenvectors corresponding to the \( A_R \) (\( A_T \)) smallest eigenvalues of \( C_{EI} \) (\( \hat{C}_{EI} \)), i.e.,

\[
\hat{P}_R = \left[ u_{N_R - A_R + 1} \left( C_{EI} \right), u_{N_R - A_R + 2} \left( C_{EI} \right), \ldots, u_{N_R} \left( C_{EI} \right) \right]
\]

\[
\hat{P}_T = \left[ u_{N_T - A_T + 1} \left( \hat{C}_{EI} \right), u_{N_T - A_T + 2} \left( \hat{C}_{EI} \right), \ldots, u_{N_T} \left( \hat{C}_{EI} \right) \right]
\]

(19)

In \cite{[19]}, \( A_R \) and \( A_T \) can be viewed as parameters to balance the achievable EI power and available degree of freedom (DOF) for data transmission, which should be optimized with respect to specific metric. This will be considered in section VI.

**Inner BF Matrix:** The inner BF matrices \( W_R \in \mathbb{C}^{A_R \times K} \) and \( W_T \in \mathbb{C}^{A_T \times K} \) are designed to realize the multi-user communication. \( W_R \) and \( W_T \) can be designed with different criteria, e.g., maximizing the desired signal power which corresponding to the eigen BF, or minimizing the MUI which corresponding to the zero-forcing (ZF) scheme \cite{[1, 2, 3]}. We adopt the latter one since the ZF scheme is known to approach the asymptotic limit (in \( \min \{N_R, N_T\} \)) of achievable rate faster as the number of relay antennas increases \cite{[9], [10]}. For given BF vectors at sources and destinations, define effective channel vectors of \( S_k \rightarrow R \) and \( R \rightarrow D_k \) channels as \( \hat{H}_{SR,k} = B_R H_{SR,k} P_{SK} \) and \( \hat{H}_{RD,k} = B_T H_{RD,k} P_{DK} \), and let \( \hat{H}_{SR} = [\hat{h}_{SR,1}, \ldots, \hat{h}_{SR,K}] \) and \( \hat{H}_{RD} = [\hat{h}_{RD,1}, \ldots, \hat{h}_{RD,K}] \), the inner BF matrices can be written as

\[
W_R = H_{SR} \left( H_{SR}^H H_{SR} \right)^{-1} \hat{H}_{SR,1} \hat{H}_{SR,1}^H \hat{H}_{SR}
\]

(20)

\[
W_T = H_{RD} \left( H_{RD}^H H_{RD} \right)^{-1} \hat{H}_{RD,1} \hat{H}_{RD,1}^H \hat{H}_{RD}
\]

where \( \Upsilon \) is a diagonal normalized matrix with \( \Upsilon_{i,i} = \left( \hat{h}_{R,R} \right)_{i,i} \). In practice, \( H_{SR,k} \) and \( H_{RD,k} \) should be estimated in order to compute \( 20 \). This will be considered in section V-B. Note that \cite{[20]} requires \( K \leq \min \{N_R, N_T\} \) so that \( A_R \) and \( A_T \) can be selected to ensure the invertibility of \( \hat{H}_{SR} \hat{H}_{SR}^H \) and \( \hat{H}_{RD} \hat{H}_{RD}^H \).

2) **Transceiver Design at the Sources and Destinations:** Similar to \cite{[12], [17]}, we assume no instantaneous knowledge of channels at sources and destinations. This is reasonable since the amount of feedback could be very huge and unaffordable in MM-FDR. However, it is assumed that the local statistical knowledge of channels (i.e., \( C_{SR,k}, \hat{C}_{SR,k} \) for \( S_k \) and \( C_{RD,k}, \hat{C}_{RD,k} \) for \( D_k \)) can be obtained. As observed from \cite{[15]}, when \( N_S = N_D = 1 \), the achievable rates of \( S_k \rightarrow R \) channel and \( R \rightarrow D_k \) channel are limited by the transmit distortion noise at \( S_k \) and receive distortion noise at \( D_k \). This motivates us to design \( p_{Sk} \) and \( p_{Dk} \) to suppress these negative factors with the antenna arrays of \( S_k \) and \( D_k \).

**Design of \( p_{Sk} \):** Intuitively, according to \cite{[3]} we can design the following problem

\[
\hat{P}_{Sk} = \arg \max_{\|p_{Sk}\| = 1} \mathbb{E} \left[ \Tr \left( H_{SR,k} p_{Sk}^H \hat{H}_{SR,k}^H \right) \right]
\]

\[
\hat{p}_{Sk} = \arg \max_{\|p_{Sk}\| = 1} \mathbb{E} \left[ \Tr \left( H_{SR,k} \Theta_{Sk}^T \hat{H}_{SR,k}^H \right) \right]
\]

\[
= \arg \max_{\|p_{Sk}\| = 1} \frac{1}{\nu_{Sk}} \Tr \left( \hat{C}_{SR,k} p_{Sk} \right)
\]

(21)

where \( \Theta_{Sk}^T \) is the covariance matrix of \( t_{Sk} \), i.e., \( \Theta_{Sk}^T = \mathbb{E} [t_{Sk} \mathbb{E} [t_{Sk}^*]] = \nu_{Sk} \text{diag} (p_{Sk}^H p_{Sk}) \). The second step follows from a similar derivation with that in \cite{[55]} and the
last step is based on assumption A2. From (19), the right-hand side of (21) can be interpreted as the average signal to transmit distortion noise ratio at $S_k$. The solution of problem (21) is

$$\hat{p}_{S,k} = u_1 \left( \tilde{C}_{SR,k} \right)$$  \hspace{1cm} (22)

**Design of $p_{D,k}$:** Similarly, based on (3), the received BF vector at $D_k$ can be derived by solving the problem

$$\hat{p}_{D,k} = \arg \max_{\|p_{D,k}\|=1} \mathbb{E} \left[ \exp \left\{ \text{Tr} \left[ (\hat{h}_{RD,k} p_{D,k} \hat{h}_{RD,k}^H) \right] \right\} \right],$$

which results in

$$\hat{p}_{D,k} = u_1 \left( \tilde{C}_{RD,k} \right)$$  \hspace{1cm} (23)

**B. Reduced Dimension Channel Estimation**

During a training phase, each source/destination transmits pilot sequence sequentially which allows the relay to compute the estimates of channels. With HIA scheme, it is sufficient to estimate the effective channel vectors $\hat{h}_{SR,k}$ and $\hat{h}_{RD,k}$. This allows a reduced dimension estimation scheme. For brevity, we consider the estimation of $\hat{h}_{SR,k}$. Let $\phi \in \mathbb{C}^{N_S \times 1}$ and $E_T$ denote the pilot sequence and power of each pilot symbol. $\phi$ is multiplied by $p_{S,k}$ and transmitted by $S_k$. The received pilot matrix at the relay is expressed as

$$Z_{SR,k} = H_{SR,k} (p_{S,k} \phi^T + T_{SR,k}) + R_{SR,k} + N_{SR,k}$$  \hspace{1cm} (24)

where $T_{SR,k} \in \mathbb{C}^{N_S \times T}$ denotes the transmit distortion noise of $S_k$, whose $l$th column has distribution based on model (1) $\mathcal{CN}(0, \mu_{S,l} E_T \text{diag}(\hat{p}_{S,k} \hat{p}_{S,k}^H))$. $R_{SR,k} \in \mathbb{C}^{N_S \times T}$ denotes receive distortion noise at the relay. With model (2), the $l$th column of $R_{SR,k}$ has distribution $\mathcal{CN}(0, \mu_{R,l} \text{diag}(\mathbb{E}[Z_{SR,k}(Z_{SR,k} - H_{SR,k} \phi \phi^H)]))$. $N_{SR,k}$ is the AWGN matrix, whose elements are i.i.d and distributed as $\mathcal{CN}(0,1)$. By multiplying both side of (24) with $\hat{p}_{R,k}^H$, we have

$$\tilde{Z}_{SR,k} = \hat{h}_{SR,k} \phi^T + \hat{p}_{R,k}^H H_{SR,k} T_{SR,k} + \hat{p}_{R,k}^H R_{SR,k} + \hat{p}_{R,k}^H N_{SR,k}$$  \hspace{1cm} (25)

From (25), as a merit of HIA scheme, the total length of pilot sequences to obtain the estimates of all effective channels can be as less as $2K$ (when $\tau$ is set to 1).

**Theorem 3:** The LMMSE estimator of effective channel $h_{SR,k}$ can be expressed as

$$\hat{h}_{SR,k} = C_{SR,k} \Gamma_{SR,k} \tilde{Z}_{SR,k}$$  \hspace{1cm} (26)

where $C_{SR,k}$ is the covariance matrix of $h_{SR,k}$, which can be expressed as $C_{SR,k} = \beta_{SR,k} \lambda_1 (C_{SR,k}) \hat{p}_{R,k} \gamma_{SR,k}$. $\gamma_{SR,k}$ is given by $\gamma_{SR,k} = \hat{h}_{SR,k} + \frac{1}{\tau E_T} \hat{p}_{R,k}^H H_{SR,k} T_{SR,k} \phi^T + \frac{1}{\tau E_T} \hat{p}_{R,k}^H R_{SR,k} \phi^T + \frac{1}{\tau E_T} \hat{p}_{R,k}^H N_{SR,k} \phi^T$ and

$$\Gamma_{SR,k} = \left( C_{SR,k} + \frac{\nu_{SR,k}}{\lambda_1 (C_{SR,k})} C_{SR,k} + \frac{1}{\tau E_T} I_{AR} \right)^{-1} + \frac{\mu_R}{\tau} \left( \frac{1}{E_T} + \beta_{SR,k} \left( \lambda_1 (C_{SR,k}) + \nu_{SR,k} \right) \right) I_{AR}$$  \hspace{1cm} (27)

The real effective channel $h_{SR,k}$ can be decomposed as $h_{SR,k} = \tilde{h}_{SR,k} + \Delta h_{SR,k}$ with $\Delta h_{SR,k}$ denoting the estimation error. $\tilde{h}_{SR,k}$ and $\Delta h_{SR,k}$ are uncorrelated whose covariance matrices can be expressed as $\tilde{C}_{SR,k} = C_{SR,k} \Gamma_{SR,k} C_{SR,k}$ and $C_{SR,k} \Gamma_{SR,k} - C_{SR,k}$, respectively.

**Proof:** A sketch of the proof is presented in Appendix-D.

Different from that with ideal hardware [23], the LMMSE estimator is not equivalent to MMSE estimator, since the received pilot signal is corrupted by the term $\hat{P}_R^H h_{SR,k} T_{SR,k}$ (due to transmit imperfection of $S_k$), which is not independent with the channel to be estimated. There might exist non-linear estimator that results in smaller MSE. However, the difference should be small since the distortion noises are relatively weak.

The LMMSE estimator of $h_{RD,k}$ can be obtained using the same approach. For further analysis, we define $\tilde{C}_{RD,k}$ and $\tilde{C}_{RD,k}$ as the covariance matrix and LMMSE estimates of $h_{RD,k}$, respectively. Moreover, we define $\Gamma_{RD,k} = \tilde{C}_{RD,k} \tilde{C}_{RD,k}$ as the covariance matrix of $\tilde{C}_{RD,k}$, where $\Gamma_{RD,k}$ is given by

$$\Gamma_{RD,k} = \left( C_{RD,k} + \frac{\nu_{RD,k}}{\lambda_1 (C_{RD,k})} C_{RD,k} + \frac{1}{\tau E_T} I_{AR} \right)^{-1} + \frac{\mu_R}{\tau} \left( \frac{1}{E_T} + \beta_{RD,k} \left( \lambda_1 (C_{RD,k}) + \nu_{RD,k} \right) \right) I_{AR}$$  \hspace{1cm} (28)

**C. Achievable Rate Analysis**

This subsection analyzes the achievable rate of HIA-MMFDR. The achievable rate expressions are derived based on the bounding technique in [34].

1) **Achievable Rate of $S_k \rightarrow R$ Channel:** By treating $\sqrt{E_{SR,k} \mathbb{E} [w_{SR,k}^H h_{SR,k} p_{S,k}^H s_{k}[u]]}$ as the desired signal at $S_k \rightarrow R$ channel, and approximating the remaining terms in (4), i.e., $\tilde{y}_{SR}[u] - \sqrt{E_{SR,k} \mathbb{E} [w_{SR,k}^H h_{SR,k} p_{S,k}^H s_{k}[u]]}$, using the worst-case uncorrelated additive Gaussian noise with the same variance, the rate in (29) is achievable on $S_k \rightarrow R$ channel, where $E_{R_k}$, $MUI_{R_k}$, $D_{R_k}$ and $D_{R_k}$ denote the EI, MUI, effective distortion due to transmit imperfection of sources and effective distortion due to receive imperfection of relay respectively, which are given by

$$E_{R_k} = w_{R_k}^H H_{EI} \left( W_T \Lambda R W_T^H + \mathbb{E} [T_{SR}[u] t_{SR}^H[u]] \right) H_{EI}^T w_{R_k}$$

$$MUI_{R_k} = \sum_{j=1 \neq k} E_{S,j} \left| w_{R_k}^H H_{SR,j} \hat{p}_{S,j} \right|^2$$

$$D_{R_k} = \sum_{j=1}^K \left| w_{R_k}^H H_{SR,j} \mathbb{E} [t_{SR}[u] t_{SR}^H[u]] \right| H_{SR,j}^H w_{R_k}$$

$$D_{R_k} = \sum_{j=1}^K \left| H_{SR,j} \mathbb{E} [t_{SR}[u] t_{SR}^H[u]] \right| H_{SR,j}^H w_{R_k}$$

$$D_{R_k} = \sum_{j=1}^K \left| H_{SR,j} \mathbb{E} [t_{SR}[u] t_{SR}^H[u]] \right| H_{SR,j}^H w_{R_k}$$

Wherein, the BF matrices at the relay are given by

$$W_R = \hat{P}_R^H H_{SR} \left( \hat{H}_{SR}^H \hat{H}_{SR} \right)^{-1} \hat{Y}^{-1/2}$$

$$W_T = \hat{P}_T^H H_{RD} \left( \hat{H}_{RD}^H \hat{H}_{RD} \right)^{-1} \hat{Y}^{-1/2}$$  \hspace{1cm} (31)
\[ \mathbf{R}_{\text{HIA}} = \frac{T - 2K\tau}{T} \log_2 \left( 1 + \frac{E_{S,k} \mathbb{E} \left[ w_{R,k}^H \mathbf{H}_{SR,k} \mathbf{p}_{S,k} \right]^2}{E_{S,k} \mathbb{E} \left[ w_{R,k}^H \mathbf{H}_{SR,k} \mathbf{p}_{S,k} \right] + E_{E,k} + E_{MUI,k} + E_{D,k} + E_{D,k}^R + E_{\|w_{R,k}^H\|^2}} \right) \]  

where \( \mathbf{H}_i = [\mathbf{h}_{i,1}, \cdots, \mathbf{h}_{i,k}] (i \in \{SR, RD\}) \) and \( \mathbf{Y} \) is a diagonal normalized matrix with \( |\mathbf{Y}|_{ii} = e_{ii} \left( \mathbf{H}_{RD}^H \mathbf{H}_{RD} \right)^{-1} \). Note that we have replaced the effective channels in inner BF matrices with their estimates obtained in Theorem 3.

\[ \mathbf{R}_{\text{HIA}} = \frac{T - 2K\tau}{T} \log_2 \left( 1 + \frac{E_{S,k} \mathbb{E} \left[ w_{R,k}^H \mathbf{H}_{SR,k} \mathbf{p}_{S,k} \right]^2}{E_{S,k} \mathbb{E} \left[ w_{R,k}^H \mathbf{H}_{SR,k} \mathbf{p}_{S,k} \right] + E_{E,k} + E_{MUI,k} + E_{D,k} + E_{D,k}^R + E_{\|w_{R,k}^H\|^2}} \right) \]  

\[ \text{Effect of EI:} \] With the proposed HIA scheme, the scaling behavior of EI is high correlated with \( \lambda_{N_R - A_R + 1} \) (1) and \( \lambda_{N_T - A_T + 1} \) (1). With the assumption A1, we have \( \lambda_{N_R - A_R + 1} \leq O(1) \) and \( \lambda_{N_T - A_T + 1} \leq O(1) \). In fact, under specific spatial correlation model (e.g., the physical channel model with a fixed number of angular bins (11)), it is possible to make \( \lambda_{N_R - A_R + 1} < O(1) \) and \( \lambda_{N_T - A_T + 1} < O(1) \) by selecting proper \( A_R \) and \( A_T \). In this case, the power of EI can decrease faster than \( O(K N^{-1}_R) \) in the large-\( (N_R, N_T) \) regime. Note that this is in contrast with the result in [12], which shown that the power of EI decreases exactly with \( O(K N^{-1}_R) \).

\[ \text{Effect of MUI:} \] The scaling behavior of MUI confirms well with the classic result in MM-FDR with perfect hardware [12], i.e., the MUI diminishes as \( N_R \to \infty \) if \( K \ll N_R \).

\[ \text{Effect of Transmit Imperfection of Sources:} \] (i) When \( N_S \) is fixed, an interesting observation is that the term \( E_{S,k} \mathbb{E} \left[ w_{R,k}^H \mathbf{H}_{SR,k} \mathbf{p}_{S,k} \right] \), which is widely considered \( O(N^{-1}_R) \) terms with perfect hardware, scales as \( O(1) \). As a result, the achievable rate of \( S_k \to R \) channel is limited by the joint effect of \( E_{S,k} \mathbb{E} \left[ w_{R,k}^H \mathbf{H}_{SR,k} \mathbf{p}_{S,k} \right] \) and the effective distortion due to the transmit imperfection of sources, i.e., \( \mathbf{D}_{R,k}^T \). In fact, the variation of scaling behavior of \( E_{S,k} \mathbb{E} \left[ w_{R,k}^H \mathbf{H}_{SR,k} \mathbf{p}_{S,k} \right] \) can also be viewed as a “negative effect” of transmit imperfection of sources. To see this, inserting the expression of \( \delta_{S,R}^T \) (given by Lemma 1 in Appendix-A) into (12) and letting \( N_R \to \infty \), we can obtain \( E_{S,k} \mathbb{E} \left[ w_{R,k}^H \mathbf{H}_{SR,k} \mathbf{p}_{S,k} \right] = \frac{\nu_{S,k}E_{S,k}g_{kk} + \sum_{j=1}^{K} \nu_{S,j}E_{S,j}g_{kj}}{\frac{\mu_R}{\mathbf{C}_{SR,k}} + \sum_{j=1}^{K} E_{S,j} \beta_{S,R,j} \mathbf{C}_{SR,j} + \nu_{S,R} \mathbf{C}_{SR,R} + 1} \) (32), where \( \delta_{S,R}^T \) is defined as \( \delta_{S,R}^T = E[\Delta \mathbf{H}_{SR,k} \mathbf{H}_{SR,k}^H] \) and is given by Lemma 1 in Appendix-A. \( g_{kk} \) and \( g_{kj} \) are given by (61) and (62) in the Appendix-E. \( \Omega_{S,k} \) and \( \Omega_R \) are expressed as

\[ \Omega_{S,k} = \mathbf{p}_{S,k} \mathbf{H}_{S,k} + \nu_{S,k} \mathbf{p}_{S,k} \mathbf{d}_{S,k} \]  

\[ \Omega_R = \sum_{l=1}^{K} E_{R,l} \mathbf{C}_{RD,l} \mathbf{p}_{S,k} \mathbf{d}_{S,k} + \nu_{R} \mathbf{p}_{S,k} \mathbf{d}_{S,k} \mathbf{d}_{S,k}^H \]  

The scaling behaviors of above factors are shown in Table 1.

\[ \text{Proof:} \] See Appendix-E.

Some important results on the effect of EI, MUI, and hardware impairments on achievable rate of \( S_k \to R \) channel can be obtained from Theorem 4 and Table 1.

1] By increasing \( A_R \), this can always be achieved since \( \mathbf{C}_{SR,k} \) approaches \( N_R \) as \( A_R \) increases from assumption A2.
TABLE I: Scaling behaviors of different factors in Theorem 4

| Desired Signal | AWGN | $E_{s,j}$ var $[w_{j,R}^n \mathbf{H}_{D,R,k}]$ | $\mathbb{E} [\text{EL}_j]$ | $\mathbb{E} [\text{MUI}_{D,k}]$ | $\mathbb{E} [\text{D}^R_{D,k}]$ | $\mathbb{E} [\text{D}^R_{D,k}]$ |
|----------------|-------|---------------------------------|----------------|----------------|----------------|----------------|
| Perfect Hardware | $O(1)$ | $O(N_j)$ | $O(N_j^2)$ | $O(N_j^2)$ | $O(N_j)$ | $O(1)$ |
| Imperfect Hardware | $O(1)$ | $O(N_j)$ | $O(N_j)$ | $O(N_j^2)$ | $O(1)$ | $O(1)$ |

TABLE II: Scaling behaviors of different factors in Theorem 5

| Desired Signal | AWGN | $E_{s,j}$ var $[p_{D,k,R}^n \mathbf{H}_{R,D,k}]$ | $\mathbb{E} [\text{MUL}_{D,k}]$ | $\mathbb{E} [\text{D}^R_{D,k}]$ |
|----------------|-------|---------------------------------|----------------|----------------|
| Perfect Hardware | $O(N_j)$ | $O(1)$ | $O(1)$ | $O(1)$ |
| Imperfect Hardware | $O(1)$ | $O(1)$ | $O(N_j)$ | $O(N_j)$ |

distortion due to transmit imperfection of relay and effective distortion due to receive imperfection of destinations, respectively, which are given by

$$
\text{MU}_{D,R,k} = \sum_{j=1,j \neq k}^{K} E_{R,j} \left| \mathbf{p}_{D,k}^H \mathbf{H}_{R,D,k} \mathbf{w}_{T,j} \right|^2
$$

$$
\text{D}^R_{D,k} = \mathbf{p}_{D,k}^H \mathbf{H}_{R,D,k} \mathbf{w}_{T,k} \mathbf{p}_{D,k}^T
$$

Theorem 5: Assume that assumptions A1 and A2 hold, and $A_T$ is selected so that $\text{Tr} \left( \mathbf{C}_{R,D,k} \right) = O(N_T)$. With HIA scheme, the achievable rate of $R \to D_k$ channel in the large-$(N_R,N_T)$ regime is given by (34), where the power of desired signal is $E_{R,k} \left| \mathbb{E} \left[ \mathbf{p}_{D,k}^H \mathbf{H}_{R,D,k} \mathbf{w}_{T,k} \right] \right|^2 = E_{R,k} \text{Tr} \left( \mathbf{C}_{R,D,k} \right)$, and

$$
E_{R,k} \mathbb{E} \left[ \text{MUI}_{D,k} \right] = \sum_{j=1,j \neq k}^{K} E_{R,j} \text{Tr} \left( \mathbf{C}_{R,D,k} \right) \left( \mathbf{C}_{D,R,k} - \mathbf{C}_{R,D,k} \right) \left( \mathbf{C}_{R,D,k} \right)^{-1} E_{R,k} \delta_{D,R,k}^R
$$

$$
E_{R,k} \mathbb{E} \left[ \text{D}^R_{D,k} \right] = \mathbf{C}_{R,D,k} \mathbf{C}_{R,D,k} \mathbf{C}_{R,D,k} \mathbf{C}_{R,D,k} \mathbf{C}_{R,D,k}
$$

Proof: The proof is similar with that for Theorem 4 and thus it is neglected.

From Theorem 5 and Table III one can observe a strong similarity between the effects of hardware impairments on $R \to D_k$ channel and that on $S_k \to R$ channel. In particular, the effective distortion due to the transmit imperfection of relay $\text{D}^R_{D,k}$ is suppressed by the large transmit array of the relay, and it diminishes as $N_T \to \infty$ and $K \ll N_T$. Moreover, when $N_D$ is fixed, the achievable rate of $R \to D_k$ channel is limited by the term $E_{R,k} \mathbb{E} \left[ \mathbf{p}_{D,k}^H \mathbf{H}_{R,D,k} \mathbf{w}_{T,k} \right]$ and distortion noise caused by receive imperfection of $D_k$. However, as $N_D$ scales with $N_T$ and $K \ll N_T$, these factors are suppressed by $\frac{1}{N_D}$. This is because that applying $\mathbf{p}_{D,k}$ can be viewed as a post-suppression on receive distortion noise.

With HIA scheme, the following end-to-end rate for $S_k \to R \to D_k$ channel is achievable in the large-$(N_R,N_T)$ regime

$$
R_k^{\text{HIA}} = \min \left( \mathbf{R}_{S,R,k}, \mathbf{R}_{D,R,k}^{\text{HIA}} \right)
$$

Different from that for upper bound in the last section, $R_k^{\text{HIA}}$ cannot be expressed in a simple form like (13). However, based on Table I and Table III one can simply deduce that $R_k^{\text{Lower}}$ will converge to a finite ceiling if $N_S$ and $N_D$ are fixed. However, with HIA scheme, the end-to-end achievable rate grows without bound as $\min(N_R,N_T) \to \infty$ if $N_S$ and $N_D$ scale with $\min(N_R,N_T)$ and $K \ll \min(N_R,N_T)$. This is the same to the situation with ideal hardware [12].

D. Discussion on Hardware Design

Based on the achievable rate expressions, we discuss the hardware design of sources, destinations and relay in this subsection. Although the achievable rate expressions hold for arbitrary $K \leq \min(N_R,N_T)$, we will restrict our analysis to $K \approx \min(N_R,N_T)$.

In MM-FDR with hardware impairments, it is also essential to control the number of S-D pairs (by, e.g., some user scheduling scheme) to suppress the EI and distortions.
1) Hardware Design of Sources and Destinations: As $N_S$ and $N_D$ are fixed and small (due to e.g., size limitation), from Theorem 4 and 5, it is of great importance to reduce $\nu_{S,k}, \nu_{D,k}$ and $\mu_{D,k}$ in order to alleviate the “ceiling effect” on achievable rate. This indicates that it is beneficial to use high-quality hardware at sources and destinations, since the increased distortion noises due to hardware impairments at sources and destinations are lower order terms when compared with desired signal powers. This means that the sources and destinations can decrease the hardware quality without hurting the performance greatly.

Example: To get a clear insight, we consider the achievable rate of $S \rightarrow R$ channel given by Theorem 4. We assume that there is only one S-D pair and let $\nu_R = \mu_R$. When $N_S$ scales with $N_R$, the achievable rate of $S_k \rightarrow R$ channel reduces to

$$\begin{align*}
R^{HIA}_{SR,k} = & \frac{T - 2K\tau}{T} \times \log_2 \left( 1 + \frac{1}{c_1 \nu_{S,1}N_R^{-1} + c_2 \mu_R N_R^{-1} + O(N_R^{-1})} \right), \\
& \text{where} \ c_1 \text{ and } c_2 \text{ are positive constants independent of } \nu_{S,1}, \mu_R \text{ and } \{N_S, \ N_R\}. \text{ In} \ [35], \text{ we have replaced } N_S \text{ with the product of } N_R \text{ and some constant. Assume we wish to achieve rate } R_T. \text{ According to} \ [35], \text{ we have} (N_R \rightarrow \infty) \ c_1 \nu_{S,1} + c_2 \mu_R = \left( 2 \tau - \frac{1}{2} \frac{N_R}{\tau^2} \right) - 1 - O(N_R^{-1})N_R \\
& \approx \left( 2 \tau - \frac{1}{2} \frac{N_R}{\tau^2} \right) - 1 \ N_R. \text{ The result is encouraging since it implies that we can increase } \nu_{S,1} \text{ and } \mu_{R} \text{ linearly as } N_R \text{ increase (with } N_S \text{ scaling with } N_R) \text{ without degrading the achievable rate. As a result, inexpensive MM-FDR is possible.}
\end{align*}$$

VI. JOINT DOF AND POWER OPTIMIZATION

In this section, we propose a low-complexity JDPO algorithm to maximize the SE (defined as the sum of all destinations’ achievable rates) of HIA-MM-FDR, subject to the maximum power constrains. The achievable rate expressions obtained in section V-C are utilized in the proposed algorithm. The algorithm needs only statistical knowledge of channels. Therefore, it can be computed offline at a central node (e.g., relay) and then broadcasts to the other nodes.

Let $E_{S,k}^{\max}$ and $E_{R,k}^{\max}$ be the maximum transmit power constraints at $S_k$ and relay, respectively, the SE optimization problem can be formulated as follows

$$\begin{align*}
\max_{E_{S,1}, \cdots, E_{S,K}; E_{R,1}, \cdots, E_{R,K}; A_S; A_T} \quad & SE = \sum_{k=1}^{K} \min \left\{ R^{HIA}_{SR,k}, R^{HIA}_{RD,k} \right\} \\
\text{s.t.} & \quad 0 \leq E_{S,k} \leq E_{S,k}^{\max}, k = 1, \cdots, K \\
& \quad \sum_{k=1}^{K} E_{R,k} \leq E_{R}^{\max} \\
& \quad A_R \in A_R, A_T \in A_T
\end{align*}$$

where $A_R = \{ K, K + 1, \cdots, N_R \}$ and $A_T = \{ K, K + 1, \cdots, N_T \}$.

According to Theorem 4 and Theorem 5, we rewrite $R^{HIA}_{SR,k}$ and $R^{HIA}_{RD,k}$ as

$$R^{HIA}_{SR,k} = \frac{T - 2K\tau}{T} \log_2 \left( (1 + \gamma_{SR,k}) \right) \text{ and } R^{HIA}_{RD,k} = \frac{T - 2K\tau}{T} \log_2 \left( (1 + \gamma_{RD,k}) \right),$$

where $\gamma_{SR,k}$ and $\gamma_{RD,k}$ denote the effective received SINRs of $S_k \rightarrow R$ and $R \rightarrow D_k$ channels respectively, which can be expressed as

$$\begin{align*}
\gamma_{SR,k} = & \frac{E_{S,k}}{\sum_{j=1}^{K} a_{kj} E_{S,j} + \sum_{j=1}^{K} b_{kj} E_{R,j} + \frac{(\mu_R + 1)}{R_T(c_{SR,k})}}, \\
\gamma_{RD,k} = & \frac{\text{Tr} \left( \hat{C}_{RD,k} \right) E_{R,k}}{\sum_{j=1}^{K} b_{kj} E_{R,j} + \mu_D + 1} \\
& \text{where} \ a_{kj} = \frac{\delta_{SR,k} + \nu_{S,k} \nu_{j} + \mu_R \Omega_{S,j}}{\text{Tr} \left( \hat{C}_{SR,k} \right)}, \\
& \text{and} \ b_{kj} = \frac{\beta_{EI} \text{Tr} \left( \hat{C}_{EI} \hat{P}_T \hat{C}_{RD,j} \hat{P}_T^H + \nu_R \text{diag} \left( \hat{P}_T \hat{C}_{RD,j} \hat{P}_T^H \right) \right)}{\text{Tr} \left( \hat{C}_{SR,k} \right) \text{Tr} \left( \hat{C}_{RD,j} \right)}.
\end{align*}$$

\footnote{We consider the achievable rate of $S_k \rightarrow R$ channel to simplify the analysis. Similar result can be obtained based on the achievable rate of $R \rightarrow D$ channel in Theorem 5. Therefore, the analysis is also valid for end-to-end achievable rate.}
as $\gamma_k = \min\{\gamma_{SR,k}, \gamma_{RD,k}\}$. (39) can be rewritten as

$$\min_{E_{S,1},\ldots,E_{S,K};E_{R,1},\ldots,E_{R,K};A_{R},A_{T}}\prod_{k=1}^K (1 + \gamma_k)^{-1} \sum_{j=1}^K a_{R,k}^j E_{S,j} + \sum_{j=1}^K b_{R,k}^j E_{R,j} + (\mu + 1) \left(\text{Tr}\left(\mathbf{C}_{SR,k}\right)\right)^{-1} \geq \gamma_k,$$

$s.t.$

$$C_1: \sum_{j=1}^K a_{R,k}^j E_{S,j} \leq E_{S,k}, k = 1, 2, \ldots, K$$

$$C_2: \sum_{j=1}^K b_{R,k}^j E_{R,j} \leq E_{R,k}$$

$$C_3: A_{R} \in A_{R}, A_{T} \in A_{T}$$

By using some algebraic manipulations on the inequality constraints, $C_1, C_2$ can be rewritten as

$$C_1: \sum_{j=1}^K a_{R,k}^j E_{S,j} \leq E_{S,k}$$

$$C_2: \sum_{j=1}^K b_{R,k}^j E_{R,j} \leq E_{R,k}$$

Problem (41) is a combinatorial optimization problem which is in general NP hard. To solve (41), we propose a JDPO algorithm to find suboptimal solution. Our strategy is as follows. First, using the similar approach as that in [35], we present a heuristic approach to find suboptimal solution. In this paper, we focus on the target function does not change dramatically as $\gamma_k$ is quite large. For example, using Algorithm 2. The algorithm converges to a local optimum since the target function is improved in each iteration. The complexity of Algorithm 2 is upper bounded as $L(|A_R| + |A_T|)L_{GP}C_{GP}$, where $C_{GP}$ is the complexity to solve GP when $A_R = N_R$ and $A_T = N_T$. Usually, GP is solved using inner point method with polynomial time. The exact expression of $C_{GP}$ is quite difficult and related with the structure of the problem. Some insights on $C_{GP}$ can be found in [37, Sec. 11.5].

Remark 2: (JDPO for EE Optimization) Instead of the SE optimization, we can also formulate an EE (defined as the ratio of the sum of the total transmit power divided by total transmit power [2]) optimization problem subject to a target $\gamma_k = \gamma_{SR,k}$. (39) can be rewritten as

$$\min_{E_{S,1},\ldots,E_{S,K};E_{R,1},\ldots,E_{R,K};A_{R},A_{T}}\prod_{k=1}^K (1 + \gamma_k)^{-1} \sum_{j=1}^K a_{R,k}^j E_{S,j} + \sum_{j=1}^K b_{R,k}^j E_{R,j} + (\mu + 1) \left(\text{Tr}\left(\mathbf{C}_{SR,k}\right)\right)^{-1} \geq \gamma_k,$$

$s.t.$

$$C_1: \sum_{j=1}^K a_{R,k}^j E_{S,j} \leq E_{S,k}, k = 1, 2, \ldots, K$$

$$C_2: \sum_{j=1}^K b_{R,k}^j E_{R,j} \leq E_{R,k}$$

$$C_3: A_{R} \in A_{R}, A_{T} \in A_{T}$$

**Algorithm 1** Solve the power control problem with fixed $A_R$ and $A_T$ by GP.

1: **Initialization**: Let $\gamma_{k(i)}$ denote the solution of $\gamma_k$ after the $i$th iteration. Compute the initial value $\gamma_{k(0)}$ using (40). Set a tolerance $\varepsilon$ and the maximum iteration times $L_{GP}$.

2: **For the $(i + 1)$th iteration:**

   - Compute $\omega_k = \gamma_{k(i)} / (1 + \gamma_{k(i)})$ and $\theta_k = (\gamma_{k(i)} - b_k) / (1 + \gamma_{k(i)})$.

   - Solve the following GP problem

     $$\min_{E_{S,1},\ldots,E_{S,K};E_{R,1},\ldots,E_{R,K};A_{R},A_{T}}\prod_{k=1}^K \theta_k^{-1} \gamma_k^{-\omega_k} s.t. C_2 \sim C_3$$

     3: If $\max_{k=1,\ldots,K} |\gamma_{k(i+1)} - \gamma_{k(i)}| < \varepsilon$ or $i + 1 = L_{GP}$, stop. Otherwise, set $i = i + 1$ and go back to step 2.

**Algorithm 2** JDPO algorithm for SE optimization.

1: **Initialization**: Set an initial $A_T$, i.e., $A_T(0)$, and a maximum repeat times $L$. Select the subsets $A_R$ and $A_T$.

2: for all $l = 1, \ldots, L$ do

3: Set $A_T = A_T(l-1)$ and compute SE using Algorithm 1 for all $A_R \in A_R$.

4: Update $A_T(l) = \arg\max_{A_R} A_T$.

5: Set $A_R = A_R(l)$ and compute SE using Algorithm 1 for all $A_T \in A_T$.

6: Update $A_T(l) = \arg\max_{A_T} A_T$.

7: end for
and $A_T$) becomes a GP. Thus, (43) can be solving by using a similar JDPO algorithm as that in Algorithm 2.

VII. SIMULATION RESULTS AND DISCUSSION

This section presents the simulation results to verify the analyses in the previous sections. Throughout this section, we set $\nu_{S,k} = \nu_S$, $\nu_{D,k} = \nu_D$, $\mu_{D,k} = \mu_D$ for convenience. The correlation matrices of desired channels are generated with the exponential correlation model [38]

$$[C_{i,k} \text{ or } \tilde{C}_{i,k} ]_{l,j} = \begin{cases} r_{i,k}^{l-j}, & l \leq j \\ (r_{i,k}^{l-j})^*, & l > j \end{cases}, i \in \{SR, RD\}$$

The model approximates the property of ULA, where the correlation between adjacent antennas is $|r_{i,k}| \in [0, 1]$ and the phase of $r_{i,k}$ describes the angle of arrival/departure as seen from the array. [39] shows how to map some of the parameters of ULA to this model. The correlation matrices of EI channel $C_{EI}$ and $\tilde{C}_{EI}$ are generated similarly with a parameter $\beta_{EI}$. For convenience, we let $|r_{i,k}| = r_{0}, \forall k \in \{1, \cdots, K\}$. The phases of $r_{i,k}$ and $\beta_{EI}$ are uniformly selected from $[0, \pi]$.

We assume that 25~35dB EI cancellation can be provided by passive EI suppression techniques (more than 40dB cancellation has been reported by using such techniques for infrastructure node [4]). The variances of EI channel (after passive cancellation) and desired channels are selected as $\beta_{EI}/\beta_{S,k} \in [0, 25]$dB. With the path loss model in [4], the above range corresponds to the setup with the distances from sources/destinations to relay varying from 250m to 500m and 10m segregation between relay transmit and receive arrays.

A. Impact of Hardware Impairments

This subsection considers the effect of hardware impairments on SE of MM-FDR. The channel coherence time is set to $T = 300$ and the length of pilot sequence is $\tau = 2$.

Fig. 1 shows the SE of MM-FDR with single antenna at sources/destinations with different levels of hardware impairments. The SEs based on transceiver scheme in section IV (which achieves the upper bound when $N_S = N_D = 1$) and HIA scheme (with $p_{S,k} = p_{D,k} = 1$) are simulated. From Fig. 1, the SE is more sensitive to hardware impairments at sources and destinations. When $\nu_S = \nu_D = \mu_D = 0.2$, the SE approaches to a finite ceiling quickly as the number of relay antennas increases. Similar results can be observed when sources and destinations are equipped with multiple but fixed number of antennas in Fig. 2. However, the result changes when $N_S$ and $N_D$ scale with $\min\{N_R, N_T\}$ and HIA scheme is used. From Fig. 2 as $N_S = \left\lceil \frac{N_R}{K} \right\rceil$ and $N_D = \left\lceil \frac{N_T}{K} \right\rceil$, similar SEs are achieved as sources/destinations or relay employ low-quality hardware, and no performance ceiling appears. This demonstrates the validity of HIA scheme. At last, it is seen that the asymptotic results in Theorem 1 & 2 and Theorem 4 & 5 match well with exact results.

As the performance is affected by impairments of both transmit and receive RF chains at the relay, we compare the effect of transmit and receive imperfections on SE in Fig. 3. We assume that the sources and destinations use high-quality hardware ($\nu_S = \nu_R = \mu_S = \mu_R = 0.01^2$). From the figure, the effect of receive imperfection is more detrimental. The reason is that, with EI at the relay, the power of distortion noise caused by receive imperfection is much stronger than that due to transmit imperfection. However, the performance difference decreases as $N_R$ increases from 0.5$N_T$ to $N_R = 2N_T$, since the power of effective receive distortion scales as $O(K N_R^{-1})$ (see Table 1). This implies that use relatively higher-quality hardware or more antennas at the receive side of relay is benefical. Using a similar setup, one can obtain a parallel conclusion for destination, i.e., the receive imperfection is more harmful than transmit imperfection. This is because that the transmit imperfection of destination only induces larger channel estimator errors, which is a part of the received signal at the destination. Thus, based on model (2), the distortion noise due to receive imperfection will be more detrimental.

![Fig. 1: SE of traditional MM-FDR with single antenna at sources/destinations, where $N_S = N_D = 1$, $K = 10$, $E_{S,k} = E_{R,k} = 8$dB, $E_T = 10$dB, $\beta_{S,k} = \beta_{R,k} = \beta_{D,k} = \beta_{EI} = 1$, $r_0 = 0.2$, $|\beta_{EI}| = 0.8$, $\nu_R = \mu_R = 0.2$, $\nu_D = \mu_D = 0.2$, $A_T = \max\{K, \frac{1}{2}N_T\}$](image1)

![Fig. 2: SE of MM-FDR with multiple antennas at sources/destinations. The setup is the same with Fig 2](image2)
Fig. 4: SE v.s. number of S-D pair \( K \), where \( \nu_S = \nu_D = \nu = 0.05, \mu_D = \mu = 0.01 \), \( K = 10 \), \( E_{S,k} = E_{R,k} = 5\text{dB}, E_T = 10\text{dB}, \beta_{SR,k} = \beta_{RD,k} = 1, \beta_{EI} = 5\text{dB}, r_0 = 0.2, |r_{EI}| = 0.8, A_R = \max\{K, \left\lfloor \frac{3N_R}{2} \right\rfloor \}, A_T = \max\{K, \left\lfloor \frac{3N_R}{2} \right\rfloor \}, N_S = \left\lfloor \frac{N_R}{K} \right\rfloor, N_D = \left\lfloor \frac{N_R}{K} \right\rfloor, T = 300, N = 2N\).

Fig. 5: Number of relay antennas required to achieve 3bit/s/Hz SE per S-D pair with different hardware qualities, where \( K = 10 \), \( \beta_{SR,k} = \beta_{RD,k} = 1, \beta_{EI} = 5\text{dB}, N_S = \left\lfloor \frac{N_R}{K} \right\rfloor, N_D = \left\lfloor \frac{N_R}{K} \right\rfloor, r_0 = 0.4, |r_{EI}| = 0.7\).

B. Impact of Number of S-D Pairs and Channel Coherent Time

Fig. 5 simulates the number of relay antennas required to achieve 3bit/s/Hz SE per S-D pair (which ideally can support the 64-QAM transmission with 1/2 channel code) with different levels of hardware quality, where we set \( \nu_S = \nu_D = \nu = 0.01 \) and \( \nu_D = \mu = \mu = 0.01 \). The figure reveals a tradeoff between the number of antennas and hardware quality, which is (as discussed in section V-D), by increasing \( N_R \)

C. Comparison with Relevant Schemes

In this subsection, the SE of HIA-MM-FDR is compared with the massive MIMO HDR (MM-HDR) \[10\] and MM-FDR with ZF-based transceiver (ZF-MM-FDR) \[12\]. The channel coherent time is set to \( T = 300 \) and the length of pilot sequence is \( \tau = 2 \). The power of pilot symbol is \( E_T = 10\text{dB} \). For HIA-MM-FDR, the maximum power constraints at sources and relay are set to \( E_{S,k}^{\max} = 5\text{dB} \) and \( E_{R,k}^{\max} = K E_{R,k}^{\max} \). Without JDPO, we set \( E_{S,k} = E_{R,k}^{\max} \), \( E_{R,k} = E_{R,k}^{\max} / K \text{dB} \). Moreover, \( A_R \) and \( A_T \) are set to \( A_R = \max\{K, \left\lfloor \frac{3N_R}{2} \right\rfloor \} \) and \( A_T = \max\{K, \left\lfloor \frac{3N_R}{2} \right\rfloor \} \). When JDPO is applied, \( E_{S,k}, E_{R,k}, A_R \) and \( A_T \) are determined by Algorithm 2. The maximum repeat time \( L \) of JDPO is set to 3 and the elements of subset \( A_R \) are picked uniformly from \( A_R \) with step \( \max\{10, \left\lfloor \frac{N_R}{K} \right\rfloor \} \). \( A_T \) is obtained with the similar approach. The above parameters are chosen so that the performance loss due to the suboptimal search approach to obtain \( A_R \) and \( A_T \) is negligible.

Fig. 5 simulates the number of relay antennas required to achieve 3bit/s/Hz SE per S-D pair (which ideally can support the 64-QAM transmission with 1/2 channel code) with different levels of hardware quality, where we set \( \nu_S = \nu_D = \nu = 0.01 \) and \( \nu_D = \mu = \mu = 0.01 \). The figure reveals a tradeoff between the number of antennas and hardware quality, which is (as discussed in section V-D), by increasing \( N_R \) and \( N_T \), we can reduce \( \nu \) and \( \mu \) linearly without degrading the SE. Meanwhile, it is seen that the proposed HIA-MM-FDR reduces the required number of relay antennas significantly when compared with ZF-MM-FDR. Moreover, the HIA-MM-FDR outperforms the MM-HDR for large \( \nu \) and \( \mu \). This is because that the distortion noises become the main limiting factor in this case when compared with EI.

The effect of asymmetric numbers of transmit and receive
where $\beta_E$ increases. By increasing $N_R$ and $N_T$, the constraint on $\beta_E$ for HIA-MM-FDR to achieve a performance gain relaxes, which indicates that HIA-MM-FDR becomes more attractive when the number of relay antennas is large. Moreover, Fig. 7 demonstrates that the proposed JDPO algorithm can reduce the constraint on $\beta_E$ significantly. In particular, as $N = 200$, a SE gain of 7.5bit/s/Hz can be achieved by HIA-MM-FDR when compared to MM-HDR as $\beta_E$ is 20dB.

Fig. 8 considers the SE-EE tradeoff of different schemes. The large-scale fading coefficients of channels are set to

$$\{\beta_{SR,1}, \ldots, \beta_{SR,K}\} = \{0.818, 0.052, 1.01, 0.026, 0.016, 0.803, 0.051, 0.383, 2.85, 0.448\}$$

$$\{\beta_{RD,1}, \ldots, \beta_{RD,K}\} = \{1.187, 0.011, 0.724, 2.11, 0.580, 0.012, 0.147, 0.085, 0.434, 0.458\}$$

which is a realization generated with the model in [12]. The EEs of HIA-MM-FDR and ZF-MM-FDR are optimized by solving the problem in Remark 2 with JDPO algorithm and [12, Algorithm 1], respectively. It is observed that HIA-MM-FDR achieves better SE-EE tradeoff when compared with ZF-MM-FDR. The gain is mainly due to the optimization of $A_R$ and $A_T$. This reason is that, if we fix $A_R = N_R$ and $A_T = N_T$, the JDPO algorithm for EE optimization is similar to [12, Algorithm 1]. The only difference is that the power for each steam at the relay, i.e., $E_{R,k}$, is optimized in JDPO algorithm and [12, Algorithm 1] optimizes only the total power of relay.

VIII. CONCLUSIONS

This paper considers the transceiver design of MM-FDR with hardware impairments. A low complexity HIA scheme is proposed to mitigate the distortion noises by exploiting the statistical knowledge of channels and antenna arrays at sources and destinations. A joint degree of freedom and power optimization algorithm is presented to further optimize the SE of HIA-MM-FDR. The analytic results demonstrate that the proposed scheme can mitigate the “ceiling effect” appears...
in traditional MM-FDR protocol, if the numbers of antennas at sources and destinations can scale with that at the relay. Moreover, simulation results show that the HIA-MM-FDR outperforms MM-FDR with traditional transceiver scheme.

**APPENDIX**

A. Useful Lemmas Related to the Channel Estimates

Lemma 1: Let \( \hat{h}_{SR,k} \) and \( \hat{h}_{RD,k} \) be the LMMSE estimates of \( h_{SR,k} \) and \( h_{RD,k} \), respectively. Define \( \delta_{SR}^k = E\left[\Delta h_{SR,k}^H \hat{h}_{SR,k}\right]^2 \) and \( \delta_{RD}^k = E\left[\Delta h_{RD,k}^H \hat{h}_{RD,k}\right]^2 \). In the large-(\( N_R, N_T \)) regime, we have

\[
\delta_{SR}^k = \frac{\nu_{SR,k}}{\tau} \left( \text{Tr} (\hat{C}_{SR,k}) \right)^2 \tilde{p}_{SR,k}^H \text{diag} (\hat{\beta}_{SR,k}) \hat{\beta}_{SR,k} \hat{p}_{SR,k} + \mu_R \frac{1}{E_T} + \beta_{SR,k} \text{Tr} (\hat{C}_{SR,k} \Omega_{S,k}) \right) \times \text{Tr} (\hat{C}_{SR,k} \Gamma_{S,k} \hat{h}_{SR,k}^H) \tag{44}
\]

The expression of \( \delta_{RD}^k \) can be obtained by replacing \( \hat{C}_{SR,k}, \hat{C}_{SR,k}, \Gamma_{SR,k}, \nu_{SR,k} \) and \( \hat{p}_{SR,k} \) in (44) with \( \hat{C}_{RD,k}, \hat{C}_{RD,k}, \Gamma_{RD,k}, \nu_{RD,k} \) and \( \hat{p}_{RD,k} \), respectively.

Proof: We show the proof for \( \delta_{SR}^k \) and the derivation for \( \delta_{RD}^k \) is similar. According to the correlation between \( \Delta h_{SR,k}^H \hat{h}_{SR,k} \) and \( \hat{h}_{SR,k} \), we have \( \delta_{SR}^k = E\left[|\Delta h_{SR,k}^H \hat{h}_{SR,k}|^2\right] = E\left[|\hat{h}_{SR,k} \hat{h}_{SR,k}|^2\right] - \left( \text{Tr} (\hat{C}_{SR,k}) \right)^2 \). With (26) and (27),

\[
E\left[|\hat{h}_{SR,k} \hat{h}_{SR,k}|^2\right] \text{can be written as}
\]

\[
E\left[|\hat{h}_{SR,k} \hat{h}_{SR,k}|^2\right] = E\left[h_{SR,k}^H C_{SR,k} \Gamma_{SR,k} h_{SR,k} \right] \tag{45}
\]

In the large-(\( N_R, N_T \)) regime, the result in [17, Lemma 4 (iv)] shows that \( E\left[|\hat{h}_{SR,k} \hat{h}_{SR,k}|^2\right] \) is given by

\[
\left( \text{Tr} (\hat{C}_{SR,k}) \right)^2 \tag{46}
\]

It can be shown that

\[
E\left[\hat{h}_{SR,k}^H C_{SR,k} \Gamma_{SR,k} \hat{p}_{R}^H \hat{H}_{SR,k} \right] \tag{47}
\]

Using the similar approach on the remaining terms in (45), Lemma 11 is obtained.

Lemma 2: Let \( C_{i,k} \) and \( \hat{C}_{i,k} \) \( (i \in \{SR, RD\}) \) be the covariance matrices of effective channel \( h_{SR,k} \) and its estimates \( \hat{h}_{SR,k} \) given by Theorem 3. If \( \text{Tr}(C_{SR,k}) = O(N_R) \) and \( \text{Tr}(C_{RD,k}) = O(N_T) \), we have \( \text{Tr}(\hat{C}_{SR,k}) = O(N_R) \) and \( \text{Tr}(\hat{C}_{RD,k}) = O(N_T) \). Moreover, \( \lim_{N_R \to \infty} \frac{\lambda_i}{N_R} > 0 \) and \( \lim_{N_T \to \infty} \frac{\lambda_i}{N_T} > 0 \) (i.e., \( A_R \) and \( A_T \) scale with \( O(N_R) \) and \( O(N_T) \), respectively).

Proof: We present the proof for \( \text{Tr}(\hat{C}_{SR,k}) \) and \( A_R \).

Using definition of \( \hat{C}_{SR,k} \) in Theorem 3 and eigenvalue decomposition \( C_{SR,k} = U \Sigma_{SR,k} U^H \), we have

\[
\text{Tr}(\hat{C}_{SR,k}) = \text{Tr}(C_{SR,k} \Gamma_{SR,k} \hat{C}_{SR,k}) \tag{48}
\]

where \( \omega_1 \) and \( \omega_2 \) are positive constants independent of \( A_R \). From (48), \( \text{Tr}(\hat{C}_{SR,k}) \) is bounded as

\[
\sum_{i=1}^{A_R} \left( \lambda_i (C_{SR,k}) \right)^2 \leq \text{Tr}(\hat{C}_{SR,k}) \leq \frac{1}{\omega_1} \text{Tr}(C_{SR,k}) \tag{49}
\]

In (49), the upper bound is an \( O(N_R) \) term. With the definition of \( C_{SR,k} \) in Theorem 3, we have

\[
\lambda_1 (C_{SR,k}) = \beta_{SR,k} \lambda_1 (C_{SR,k}) u_1^H (C_{SR,k}) \times \hat{p}_R^H C_{SR,k} \hat{p}_R u_1 (C_{SR,k}) \leq \beta_{SR,k} \lambda_1 (C_{SR,k}) \beta_{SR,k} = O(1) \tag{50}
\]

The last step is based on assumption A1. Since \( \text{Tr}(C_{SR,k}) = O(N_R) \), we can conclude that \( \lambda_1 (C_{SR,k}) = O(1) \) (otherwise \( \text{Tr}(C_{SR,k}) \leq A_R \lambda_1 (C_{SR,k}) < O(N_R) \)). Thus, the lower bound in (49) also scales with \( O(N_R) \). Moreover, based on the definition of \( C_{SR,k} \) in the Theorem 3, we have

\[
\text{Tr}(\hat{C}_{SR,k}) = \beta_{SR,k} \lambda_1 (C_{SR,k}) \sum_{i=1}^{A_R} \hat{p}_{RI}^H C_{SR,k} \hat{p}_{RI} \tag{51}
\]

where \( \hat{p}_{RI} \) is the \( l \)th column of \( \hat{P}_R \). From (51), \( A_R \) must scale \( O(N_R) \). Otherwise, \( \text{Tr}(C_{SR,k}) < O(N_R) \) based on assumption A1.
B. Proof of Theorem 7

To facilitate analysis, we first derive the covariance matrices of distortion noises $s_t[u]$ and $r_R[u]$. Based on the model \[1\], we have $\Theta_R = \text{diag}(\nu_{S,1}E_S, \ldots, \nu_{S,K}E_S)$. Moreover, define $x_S[u] = [x_{S,1}[u], \ldots, x_{S,K}[u]]^T$, $\Theta_R^R$ can be expressed based on model \[1\] as

$$
\Theta_R^R = \mu_R \text{diag} \left( \mathbb{E} \left[ H_{SR} \Omega_S \mathbf{h}_{SR}^H \right] \right) + \mu_R \text{diag} \left( \mathbb{E} \left[ H_{EI} \Omega_R \mathbf{h}_{EI}^H \right] \right) + \mu_R \mathbf{I}_{N_R}
$$

(52)

where

$$
\Omega_S = \mathbb{E} \left[ (x_S[u] + t_S[u]) (s[u] + t_S[u])^H \right] = \text{diag} \left( (1 + \nu_{S,1})E_S, \ldots, (1 + \nu_{S,K})E_S \right)
$$

and

$$
\Omega_R = \mathbb{E} \left[ (x_R[u] + t_R[u]) (x_R[u] + t_R[u])^H \right] = \mathbb{E} \left[ x_R[u] x_R^H[u] \right] + \Theta_R
$$

(53)

For further approximation, we approximate $\mathbb{E} \left[ x_R[u] x_R^H[u] \right]$ in \ref{53} with $\text{diag}(\mathbb{E} \left[ x_R[u] x_R^H[u] \right]) = 1/\nu_R \Theta_R^R$. Note that the approximation results in a new upper bound on $R_{SR,k}^\text{Upper}$. (The new bound will also be referred to as $R_{SR,k}^\text{Upper}$ for convenience). Straight-forward computations yield to

$$
\mathbb{E} \left[ H_{SR} \Omega_S \mathbf{h}_{SR}^H \right] = \sum_{l=1}^{K} \nu_{S,l} E_{S,l} \mathbb{E} \left[ h_{SR,l} \mathbf{h}_{SR,l}^H \right]
$$

(54)

Following the spatial correlation model of $H_{EI}$ in section III and the approximation under \ref{53}, the expression of $\mathbb{E} \left[ H_{EI} \Omega_R \mathbf{h}_{EI}^H \right]$ can be obtained as in \ref{55}, where the third step follows from the formula vec(ABC) = $(c^T \otimes A) \text{vec}(B)$ and the fourth step is due to the fact that the elements of $X_{EI}$ are i.i.d with distribution $\mathcal{CN}(0, 1)$. Inserting \ref{54} and \ref{55} into \ref{52}, the expression of $\Theta_R^R$ in \ref{10} is obtained.

Then we derive the asymptotic bound in Theorem \ref{1} By replacing $H_{EI} \Theta_R^R H_{EI}^H$ in $Q_k$ with $\text{diag}(H_{EI} \Theta_R^R H_{EI}^H)$, we approximate $Q_k$ as $Q_k = H_{SR} \Theta_R^R H_{SR}^H + \text{diag}(H_{EI} \Theta_R^R H_{EI}^H) + \mathbf{I}_{N_R}$. The approximation results in a new upper bound on $R_{SR,k}^\text{Upper}$ since it reduces the power of residual $E_I$ after combining by $w_{SR,k}$. (The new bound will also be referred to as $R_{SR,k}^\text{Upper}$ for convenience). Based on \ref{17}, Lemma 4 (ii), we can replace $\text{diag}(H_{EI} \Theta_R^R H_{EI}^H)$ with its deterministic equivalence in the large-$N_T$ regime

$$
\frac{1}{N_T} \text{diag}(H_{EI} \Theta_R^R H_{EI}^H) = \frac{1}{N_T} \beta_{EI} \text{Tr}(C_{EI} \Theta_R^R) \mathbf{I}_{N_T}.
$$

Finally, the expression of $R_{SR,k}^\text{Upper}$ in Theorem \ref{1} can be obtained by first applying \ref{17}, Lemma 4 (ii) on $h_{SR,k}^H h_{SR,k}$, and then using [17, Theorem 1].

C. Proof of Theorem 2

Since the scenario of interest is the large-$(N_R, N_T)$ regime, we need only consider the effect of $w_{T,k}$ on $R_{SR,k}^\text{Upper}$ (given by Theorem \ref{1}), more precisely, the term $\text{Tr}(C_{EI} \Theta_R^R)$. With the assumption A2, we have $\text{diag}(C_{EI}) = N_T$. Thus, using the model \ref{1}, we have $\text{Tr}(C_{EI} \Theta_R^R) = \text{Tr}(\Theta_R^R) = \nu_R \text{Tr}(\mathbb{E} \left[ x_R[u] x_R^H[u] \right]) = \nu_R \sum_{l=1}^{K} E_{R,l}$, which is independent of $w_{T,k}$.

Then we derive the upper bound in Theorem \ref{2} Using the model \ref{4}, the power of received distortion at the relay can be derived as $E \left[ \| r_{D,k}[u] \|^2 \right] = \mu_{D,k} \left( N_T \beta_{RD,k} E_{R,k} + \nu_R N_T \beta_{RD,k} \sum_{l=1}^{K} E_{R,l} + 1 \right)$. Moreover, in the large $N_T$ regime, the term $h_{RD,k}^H \Theta_R^R h_{RD,k}$ in \ref{7} approaches $h_{RD,k}^H \Theta_R^R h_{RD,k} = \beta_{RD,k} \text{Tr}(C_{RD,k} \Theta_R^R) = \nu_R \beta_{RD,k} \sum_{l=1}^{K} E_{R,l}$ [17, Lemma 4 (ii)]. Therefore, it is sufficient to design $w_{T,k}$ to maximize the numerator of \ref{7}, which is exactly the eigen BF scheme. The resultant upper bound on achievable rate can be obtained by applying [17, Lemma 4 (ii)] and assumption A2 on the numerator of \ref{7}.

D. Proof of Theorem 3

With the expression of LMMSE estimator \ref{23}, we have $\hat{h}_{SR,k} = E \left[ h_{SR,k} \tilde{z}_{SR,k} \right] = (E \left[ \tilde{z}_{SR,k} \tilde{z}_{SR,k}^H \right])^{-1} \tilde{z}_{SR,k}$. Using the independence between $h_{SR,k}$ and distortion noises, we have

$$
\mathbb{E} \left[ h_{SR,k} \tilde{z}_{SR,k}^H \right] = \mathbb{E} \left[ \hat{H}_R^H H_{SR,k} \hat{p}_{SR,k} \hat{p}_{SR,k}^H \hat{H}_R^H \hat{h}_{SR,k} \hat{p}_R \right] = \beta_{SR,k} \lambda_1 \left( \hat{C}_{SR,k} \right) \hat{p}_R \mathbf{C}_{SR,k} \hat{p}_R \tag{56}
$$

where the second step follows from a similar derivation with that in \ref{55}. Moreover, $\mathbb{E} \left[ \tilde{z}_{SR,k} \tilde{z}_{SR,k}^H \right]$ can be expressed as

$$
\mathbb{E} \left[ \tilde{z}_{SR,k} \tilde{z}_{SR,k}^H \right] = \mathbb{E} \left[ \hat{H}_R^H H_{SR,k} \mathbf{t}_{SR,k} \phi^T \mathbf{t}_{SR,k}^T \hat{H}_R \hat{h}_{SR,k} \hat{p}_R \right] + \frac{1}{\sigma^2} \mathbb{E} \left[ \hat{H}_R^H H_{SR,k} \mathbf{t}_{SR,k} \phi^T \mathbf{t}_{SR,k}^T \hat{H}_R \hat{h}_{SR,k} \hat{p}_R \right] \tag{57}
$$

According to the independence between $h_{SR,k}$ and $T_{SR,k}$, the second term of right-hand side of \ref{57} can be rewritten as

$$
\mathbb{E} \left[ \hat{H}_R^H H_{SR,k} \mathbf{t}_{SR,k} \phi^T \mathbf{t}_{SR,k}^T \hat{H}_R \hat{h}_{SR,k} \hat{p}_R \right] = \nu_{SR,k} \tau E_T^2 \mathbb{E} \left[ \hat{H}_R^H H_{SR,k} \text{diag} \left( \hat{p}_{SR,k} \hat{p}_{SR,k}^H \right) \hat{H}_R \hat{h}_{SR,k} \hat{p}_R \right] = \nu_{SR,k} \tau E_T^2 \beta_{SR,k} \text{Tr} \left( \hat{C}_{SR,k} \text{diag} \left( \hat{p}_{SR,k} \hat{p}_{SR,k}^H \right) \right) \hat{p}_R \mathbf{C}_{SR,k} \hat{p}_R \tag{58}
$$

Using the similar approach on the remaining terms of \ref{57}, and substituting the resultant expression of \ref{57} and \ref{56} into the expression of LMMSE estimator, the result in Theorem \ref{3} is obtained. Using \ref{20}, the second order statistics of $h_{SR,k}$ and $\Delta h_{SR,k}$ in Theorem \ref{3} can be easily verified.

E. Proof of Theorem 4

We show the proof for $E[|E_{l,k}|]$ and $E[|D_{l,k}|]$ due to the space limitation. The proof for the other terms in Theorem \ref{4} is similar.

Note that $\hat{H}_R^H \hat{H}_R$ is $\nu_{SR,k} \tau E_T^2$-i.i.d. and $\hat{C}_{SR,k}$ is $\nu_{SR,k} \tau E_T^2$-i.i.d. $x_i$ on $\{SR, RD\}$. Based on Lemma \ref{2} the dimensions of
The second step is based on the property $\text{Tr}(AB) = \text{Tr}(BA)$ and independence between $\mathbf{t}_{SR,k}$ and $\mathbf{H}_{EI}$. Finally, using a similar derivation with (55), $\mathbb{E}[E_{lk}]$ in Theorem 4 is obtained.

(b) Scaling Behavior: Substituting (59) into the expression of $\mathbb{E}[E_{lk}]$ in Theorem 4 and neglecting the terms that have no effect on the scaling behavior, we have

$$
\mathbb{E}[E_{lk}] = \beta_{EI} \left( \text{Tr} \left( \mathbf{C}_{SR,k} \right) \right)^{-1} \sum_{i=1}^{K} E_{R,i} \left( \text{Tr} \left( \mathbf{C}_{RD,i} \right) \right) \times \text{Tr} \left( \mathbf{C}_{SR,k} \mathbf{H}_{SR,i} \mathbf{H}_{SR,i}^H \mathbf{R}_{SR,k} \mathbf{P}_{SR,k} \mathbf{H}_{SR,i}^H \mathbf{R}_{SR,k} \right)
$$

where the second step follows from the definition of $\mathbf{P}_{SR,k}$ and $\mathbf{P}_{SR,k}$. Based on Lemma 2 (Tr($\mathbf{C}_{SR,k}$))$^{-1} = O(N^{-1})$. Thus we can conclude that $\mathbb{E}[E_{lk}] \leq O(\lambda_{N_R-A_T+1}(\mathbf{C}_{EI}) \lambda_{N_T-A_T+1}(\mathbf{C}_{EI}) K N^{-1})$.

2) Derivation of $\mathbb{E}[D_{Rk}^T]$:

(a) Asymptotic Expression: Based on the model (1), the covariance matrix of $\mathbf{t}_{SR,k}$ can be derived as $\mathbb{E}[\mathbf{t}_{SR,k} \mathbf{t}_{SR,k}^H] = \nu_{SR} \mathbf{I}$. Thus, in the large-$N_R$, $N_T$ regime, using (58), $\mathbb{E}[D_{Rk}^T]$ can be expressed as

$$
\mathbb{E}[D_{Rk}^T] = \left( \text{Tr} \left( \mathbf{C}_{SR,k} \right) \right)^{-1} \nu_{SR} E_{SR,k} \times \text{Tr} \left( \mathbf{H}_{SR,k}^H \mathbf{H}_{SR,k} \mathbf{R}_{SR,k} \mathbf{P}_{SR,k} \mathbf{H}_{SR,k}^H \mathbf{R}_{SR,k} \right) + \left( \text{Tr} \left( \mathbf{C}_{SR,k} \right) \right)^{-1} \sum_{j=1}^{K} \nu_{SJ} E_{SJ,j} \times \text{Tr} \left( \mathbf{H}_{SR,k}^H \mathbf{H}_{SR,j} \mathbf{R}_{SR,k} \mathbf{P}_{SR,k} \mathbf{H}_{SR,j}^H \mathbf{R}_{SR,k} \right)
$$

According to the independence between $\mathbf{t}_{SR,k}$ and $\mathbf{H}_{SR,j}$ when $k \neq j$, $g_{kj}$ can be derived as

$$
g_{kj} = \beta_{SR,j} \text{Tr} \left( \mathbf{C}_{SR,k} \mathbf{P}_{SR,k} \mathbf{R}_{SR,j} \mathbf{P}_{SR} \right)
$$

Moreover, substituting (26) and (27) in Theorem 3 into (60) and using the large-$N_R$ approximation in (47), it is straight-
forward but tedious to show that $g_{kk}$ can be derived as

$$g_{kk} = \left( \text{Tr} \left( \hat{C}_{SR,k} \right) \right)^2 \hat{p}_{S,k}^H \text{diag} \left( \hat{p}_{S,k} \hat{p}_{S,k}^H \right) \hat{p}_{S,k} + \frac{\nu_{ER,k}}{\tau} \left( \text{Tr} \left( \hat{C}_{SR,k} \right) \right)^2 \left( \text{diag} \left( \hat{p}_{S,k} \hat{p}_{S,k}^H \right) \hat{C}_{SR,k} \right)^2 \right)$$

$$+ \frac{\mu_R}{\tau} \left( 1 + \beta_{SR,k} \text{Tr} \left( \hat{C}_{SR,k} \Omega_{SR,k} \right) \right) \text{Tr} \left( \hat{C}_{SR,k}^T \Gamma_{SR,k} \right)$$

$$+ \frac{1}{\tau E_T} \lambda_3 \left( \hat{C}_{SR,k} \right)$$

$$= \hat{C}_{SR,k}^T \Gamma_{SR,k}$$

(62)

where $\Omega_{SR,k} = \hat{p}_{S,k} \hat{p}_{S,k}^H + \nu_{SR,k} \text{diag} \left( \hat{p}_{S,k} \hat{p}_{S,k}^H \right)$. Inserting (61) and (62) into (60), the expression of $E[D_{R,k}^T]$ is obtained.

(b) Scaling Behavior: Using the similar approach as that for $E[E[k]]$, it can be shown that, as $N_S$ is fixed, the first and second terms on the right-hand side of (60) scale as $O(1)$ and $O(K N^{-1})$, respectively. Thus, $E[D_{R,k}^T] = O(1)$ since $K < N_R$. As $N_S$ scales with $N_R$, it is shown in [19] that the eigenvectors of $\hat{C}_{EI}$ form a unitary DFT matrix in the large $N_S$ regime. In this case, we have $\hat{d}(\hat{p}_{S,k} \hat{p}_{S,k}^H) = 1/N_S$. Applying this property and Lemma 2 on (62), it can be verified that the first term on the right-hand side of (60) scales as $O(1/N_S)$ as $N_S$ scales with $N_R$. This completes the proof.

REFERENCES

[1] T. L. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” IEEE Trans. Wireless Commun., vol. 9, no. 11, pp. 3590-3600, Nov. 2010.

[2] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, “Energy and spectral efficiency of very large multiuser MIMO systems,” IEEE Trans. Commun., vol. 61, no. 4, pp. 1436-1449, 2013.

[3] D. Bharadia, E. McMilin, and S. Katti, “Full duplex radios,” ACM Sigcomm 2013, Hong Kong, China, Aug. 2013.

[4] E. Everett, A. Sahai, and A. Sabharwal, “Passive self-interference suppression for full-duplex infrastructure nodes,” IEEE Trans. Wireless Commun., vol. 12, no. 2, pp. 680-694, 2014.

[5] H. A. Suraweera, I. Krikidis, G. Zheng, C. Yuen and P. J. Smith, “Low-complexity end-to-end performance optimization in MIMO full-duplex relay systems,” IEEE Trans. Wireless Commun., vol. 13, no. 9, pp. 3913-3927, Feb. 2014.

[6] C. Y. A. Shang, P. J. Smith, G. K. Woodward and H. A. Suraweera, “Linear transceivers for full duplex MIMO relays,” in Proc. Australian Communications Theory Workshop (AusCTW 2014), Sydney, Australia, Feb. 2014, pp. 1722.

[7] D. Nguyen, L.-N. Tran, P. Pirinen, and M. Latva-aho, “Precoding for full duplex multiuser MIMO systems: spectral and energy efficiency maximization,” IEEE Trans. Signal Process., vol. 61, no. 16, pp. 4038-4050, Aug. 2013.

[8] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan and R. Wichman, “In-band full-duplex wireless: challenges and opportunities,” IEEE J. Sel. Areas Commun., vol. 32, no. 9, pp. 1637-1652, Sep. 2014.

[9] H. Cui, L. Song, and B. Jiao, “Multi-pair two-way amplify-and-forward relaying with very large number of relay antennas,” IEEE Trans. Wireless Commun., vol. 13, no. 5, pp. 2630-2645, Aug. 2013.

[10] H. A. Suraweera, N. Hen Quoc, T. Q. Duong, and et al, “Multi-pair amplify-and-forward relaying with very large antenna arrays,” in IEEE ICC 2013, pp. 4635-4640, Jun. 2013.

[11] H. Q. Ngo, H. A. Suraweera, E. G. Larsson, “Multi-pair massive MIMO full-duplex relaying with MRC/MRT processing,” IEEE ICC 2013, pp. 4807-4813, Jun. 2013.

[12] H. Q. Ngo, H. A. Suraweera, M. Mathaihou, and E. G. Larsson, “Multi-pair full-duplex relaying with massive arrays and linear processing,” IEEE J. Sel. Areas Commun., vol. 32, no. 9, pp. 1721-1737, Sep. 2014.
[38] S. Loyka, “Channel capacity of MIMO architecture using the exponential correlation matrix,” *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 369-371, 2001.

[39] E. Björnson, D. Hammarwall, and B. Ottersten, “Exploiting quantized channel norm feedback through conditional statistics in arbitrarily correlated MIMO systems,” *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 4027-4041, 2009.

**Xiaochen Xia** received his B.E. degree in electronic science and technology from Tianjin University (TJU) in 2010, and M.S. degree in communication and information system from PLA University of Science and Technology (PLAUST), in 2013. He is currently working toward the Ph.D. degree in institution of communications engineering, PLAUST. His research interests include relaying network, full-duplex communication, Network coding, MIMO techniques. He received the 2013 excellent master degree dissertation award of Jiangsu Province, China.

**Dongmei Zhang** received the B.S. degree in institute of communications engineering, Nanjing, China in 1994, and the M.S. degree in communications and information system from the institute of communications engineering, Nanjing, China in 1997. She is currently an associate professor with the PLA University of Science and Technology, Nanjing, China. Her research interests include broadband wireless communications, cognitive radio, wireless resource allocation.

**Kui Xu** received the B.S. degree and Ph.D. degree from the PLA University of Science and Technology (PLAUST), Nanjing, China in 2004 and 2009. He is currently a lecturer in the PLAUST. His research interests include broadband wireless communications, signal processing for communications and wireless communication networks. He is the author of about 50 papers in refereed journals and conference proceedings and holds 5 patents in China. Dr. Xu currently served on the technical program committee of the IEEE WCSP 2014. He received the URSI Young Scientists Award in 2014 and the 2010 ten excellent doctor degree dissertation award of PLAUST.

**Wenfeng Ma** received the B.S. degree in microwave communication from the institute of communications engineering, Nanjing, China in 1995, and the M.S. degree in communications and information system from the institute of communications engineering, Nanjing, China in 1998, and the Ph.D. degree in communications and information system from the PLA University of Science and Technology, Nanjing, China, in 2002. He is currently an associate professor with the PLA University of Science and Technology, Nanjing, China. His research interests include wireless communication networks, broadband wireless communications.

**Youyun Xu** was graduated from Shanghai Jiao Tong University with Ph.D. Degree in Information and Communication Engineering in 1999. He is currently a professor with Nanjing Institute of Communication Engineering, China. He is also a part-time professor with the Institute of Wireless Communication Technology of Shanghai Jiao Tong University (SJTU), China. Professor Xu has more than 20 years professional experience of teaching and researching in communication theory and engineering. Now, his research interests are focusing on New Generation Wireless Mobile Communication System (IMT-Advanced and Related), Advanced Channel Coding and Modulation Techniques, Multi-user Information Theory and Radio Resource Management, Wireless Sensor Networks, Cognitive Radio Networks, etc.