Transceiver Design at Relay and Destination for Dual-Hop Non-regenerative MIMO-OFDM Relay Systems Under Channel Uncertainties

Chengwen Xing, Shaodan Ma, Yik-Chung Wu and Tung-Sang Ng

Abstract

In this paper, joint design of linear relay precoder and destination equalizer for dual-hop non-regenerative amplify-and-forward (AF) MIMO-OFDM systems under channel estimation errors is investigated. Second order moments of channel estimation errors in the two hops are first deduced. Then based on the Bayesian framework, joint design of linear robust precoder at the relay and equalizer at the destination is proposed to minimize the total mean-square-error (MSE) of the output signal at the destination. The optimal designs for both correlated and uncorrelated channel estimation errors are considered. The relationship with existing algorithms is also disclosed. Simulation results show that the proposed robust designs outperform the design based on estimated channel state information only.

Keywords: Minimum mean-square-error (MMSE), Amplify-and-forward (AF), precoder, equalizer.

Chengwen Xing, Shaodan Ma, Yik-Chung Wu and Tung-Sang Ng are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong. Phone : (852) 2857 8406, Fax : (852) 2559 8738, Email : {cwxing, sdma, ycwu, tsng }@eee.hku.hk.

*The corresponding author is Chengwen Xing.

I. Introduction

In order to enhance the coverage of base stations and quality of wireless links, dual-hop relaying is being considered to be one of the essential parts for future communication systems (e.g., LTE, IMT-Adanced, Winner Project). In dual-hop cooperative communication, relay nodes receive signal transmitted from a source and then forward it to the destination [1], [2]. Roughly speaking, there are three different relay strategies: decode-and-forward (DF), compress-and-forward (CF) and amplify-and-forward (AF). Among them, AF strategy is the most preferable for practical systems due to its low complexity [3]–[7].

On the other hand, for wideband communication, multiple-input multiple-output (MIMO) orthogonal-frequency-division-multiplexing (OFDM) has gained a lot of attention in both industrial and academic communities, due to its high spectral efficiency, spatial diversity and multiplexing gains [8]–[11]. The combination of AF and MIMO-OFDM becomes an attractive option for enabling high-speed wireless multi-media services [12].

In the last decade, linear transceiver design for various systems has been extensively investigated because of its low implementation complexity and satisfactory performance [8], [13]. For linear transceiver design, minimum mean-square-error (MMSE) is one of the most important and frequently used criteria [14]–[20]. For example, for point-to-point MIMO and MIMO-OFDM systems, linear MMSE transceiver design has been discussed in details in [14]–[16]. Linear MMSE transceiver design for multiuser MIMO systems has been considered in [17]. For single carrier AF MIMO relay systems, linear MMSE precoder at the relay and equalizer at the destination are joint designed in [19]. Furthermore, the linear MMSE transceiver design for dual hop MIMO-OFDM relay systems is proposed in [20].

In all the above works, channel state information (CSI) is assumed to be perfectly known. Unfortunately, in practical systems, CSI must be estimated and channel estimation errors are inevitable. When channel estimation errors exist, in general, two classes of robust designs can be employed: min-max and stochastic robust designs. If the distributions of channel estimation errors are known to be unbounded, stochastic robust design is preferred. Stochastic robust design includes probability-based design and Bayesian design. In this paper, we focus on Bayesian design, in which an averaged mean-square-error (MSE) performance is considered. Recently, Bayesian robust linear MMSE transceiver design under channel uncertainties has been addressed for point-to-point MIMO systems [22], [23] and point-to-point MIMO-OFDM systems [24].

In this paper, we take a step further and consider the robust linear MMSE relay precoder and destination equalizer design for dual-hop AF MIMO-OFDM relay systems without considering direct link. For channel estimation in the two hops, both the linear minimum mean square error and maximum likelihood estimators are derived, based on which the second order moments of channel estimation errors are deduced. Using the Bayesian framework, channel estimation errors are taken into account in the transceiver design criterion. Then a general closed-form solution for the optimal transceiver is proposed. Both the uncorrelated and correlated channel estimation errors are considered. The relationship between the proposed algorithm and several existing robust transceiver designs is revealed. Furthermore, simulation results demonstrate that the proposed robust algorithms provide an obvious advantage in terms of data mean-square-error (MSE) compared to the algorithm based on estimated CSI only.

This paper is organized as follows. System model is presented in Section II. Channel estimators and the corresponding covariance of channel estimation errors are derived in section III. The optimization problem for transceiver design is formulated in Section IV. In Section V, the general optimal closed-form solution for the transceiver design problem is proposed. Simulation results are given in Section VI and finally, conclusions are drawn in Section VII.

The following notations are used throughout this paper. Boldface lowercase letters denote vectors, while boldface uppercase letters denote matrices. The notations $\mathbf{Z}^{\mathrm{T}}$, $\mathbf{Z}^{\mathrm{H}}$ and $\mathbf{Z}^{*}$ denote the transpose, Hermitian and conjugate of the matrix $\mathbf{Z}$, respectively, and $\mathrm{Tr}(\mathbf{Z})$ is the trace of the matrix $\mathbf{Z}$. The symbol $\mathbf{I}_M$ denotes the $M \times M$ identity matrix, while $\mathbf{0}_{M\times N}$ denotes the $M \times N$ all zero matrix. The notation $\mathbf{Z}^{\frac{1}{2}}$ is the Hermitian square root of the positive semi-definite matrix $\mathbf{Z}$, such that $\mathbf{Z} = \mathbf{Z}^{\frac{1}{2}}\mathbf{Z}^{\frac{1}{2}}$ and $\mathbf{Z}^{\frac{1}{2}}$ is a Hermitian matrix. The symbol $\mathbb{E}\{.\}$ represents the expectation operation. The operation $\mathrm{vec}(\mathbf{Z})$ stacks the columns of the matrix $\mathbf{Z}$ into a single vector. The symbol $\otimes$ represents Kronecker product. The symbol a^{+} means $\max\{0, a\}$. The notation $\mathrm{diag}[\mathbf{A}, \mathbf{B}]$ denotes the block diagonal matrix with $\mathbf{A}$ and $\mathbf{B}$ as the diagonal elements.

II. System Model

In this paper, a dual-hop amplify-and-forward (AF) MIMO-OFDM relay cooperative communication system is considered, which consists of one source with N_S antennas, one relay with M_R receive antennas and N_R transmit antennas, and one destination with M_D antennas, as shown in Fig. 1. At the first hop, the source transmits data to the relay, and the received signal $\mathbf{x}_k$ at the relay on the k^{th} subcarrier is

$$\mathbf{x}_k = \mathbf{H}_{sr,k}\mathbf{s}_k + \mathbf{n}_{1,k} \quad k = 0, 1, \cdots K-1, \tag{1}$$

where $\mathbf{s}_k$ is the data vector transmitted by the source with covariance matrix $\mathbf{R}_{\mathbf{s}_k} = \mathbb{E}\{\mathbf{s}_k\mathbf{s}_k^{\text{H}}\}$ on the k^{th} subcarrier, and $\mathbf{R}_{\mathbf{s}_k}$ can be an arbitrary covariance matrix. The matrix $\mathbf{H}_{sr,k}$ is the MIMO channel between the source and relay on the k^{th} subcarrier. The symbol $\mathbf{n}_{1,k}$ is the additive Gaussian noise with zero mean and covariance matrix $\mathbf{R}_{n_{1,k}} = \sigma_{n_1}^2\mathbf{I}_{M_R}$ on the k^{th} subcarrier. At the relay, for each subcarrrier, the received signal $\mathbf{x}_k$ is multiplied by a precoder matrix $\mathbf{F}_k$, under a power constraint $\sum_k \text{Tr}(\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\mathbf{F}_k^{\text{H}}) \leq P_r$ where $\mathbf{R}_{\mathbf{x}_k} = \mathbb{E}\{\mathbf{x}_k\mathbf{x}_k^{\text{H}}\}$ and P_r is the maximum transmit power. Then the resulting signal is transmitted to the destination. The received data $\mathbf{y}_k$ at the destination on the k^{th} subcarrier is

$$\mathbf{y}_k = \mathbf{H}_{rd,k}\mathbf{F}_k\mathbf{H}_{sr,k}\mathbf{s}_k + \mathbf{H}_{rd,k}\mathbf{F}_k\mathbf{n}_{1,k} + \mathbf{n}_{2,k}, \tag{2}$$

where the symbol $\mathbf{n}_{2,k}$ is the additive Gaussian noise vector on the k^{th} subcarrier at the second hop with zero mean and covariance matrix $\mathbf{R}_{n_{2,k}} = \sigma_{n_2}^2\mathbf{I}_{M_D}$. In order to guarantee the transmitted data $\mathbf{s}_k$ can be recovered at the destination, it is assumed that M_R, N_R, and M_D are greater than or equal to N_S [6].

The signal $\mathbf{x}$ received at the relay and the signal $\mathbf{y}$ received at the destination in frequency domain can be compactly written as

$$\mathbf{x} = \mathbf{H}_{sr}\mathbf{s} + \mathbf{n}_1, \tag{3}$$

$$\mathbf{y} = \mathbf{H}_{rd}\mathbf{F}\mathbf{H}_{sr}\mathbf{s} + \mathbf{H}_{rd}\mathbf{F}\mathbf{n}_1 + \mathbf{n}_2, \tag{4}$$

where

$$\mathbf{y} \triangleq [\mathbf{y}_0^{\text{T}}, \cdots, \mathbf{y}_{K-1}^{\text{T}}]^{\text{T}}, \quad \mathbf{s} \triangleq [\mathbf{s}_0^{\text{T}}, \cdots, \mathbf{s}_{K-1}^{\text{T}}]^{\text{T}} \tag{5a}$$

$$\mathbf{F} \triangleq \text{diag}[\mathbf{F}_0, \cdots, \mathbf{F}_{K-1}], \tag{5b}$$

$$\mathbf{H}_{sr} \triangleq \text{diag}[\mathbf{H}_{sr,0}, \mathbf{H}_{sr,1}, \cdots, \mathbf{H}_{sr,K-1}], \tag{5c}$$

$$\mathbf{H}_{rd} \triangleq \text{diag}[\mathbf{H}_{rd,0}, \mathbf{H}_{rd,1}, \cdots, \mathbf{H}_{rd,K-1}], \tag{5d}$$

$$\mathbf{n}_1 \triangleq [\mathbf{n}_{1,0}^{\text{T}}, \mathbf{n}_{1,1}^{\text{T}}, \cdots, \mathbf{n}_{1,K-1}^{\text{T}}]^{\text{T}}, \tag{5e}$$

$$\mathbf{n}_2 \triangleq [\mathbf{n}_{2,0}^{\text{T}}, \mathbf{n}_{2,1}^{\text{T}}, \cdots, \mathbf{n}_{2,K-1}^{\text{T}}]^{\text{T}}. \tag{5f}$$

Notice that in general the matrix $\mathbf{F}$ in (4) can be an arbitrary $KN_R \times KM_R$ matrix instead of a block diagonal matrix. This corresponds to mixing the data from different subcarriers at the relay, and is referred as subcarrier cooperative AF MIMO-OFDM systems [20]. It is obvious that when the number of subcarrier K is large, transceiver design for such systems needs very high complexity. On other hand, it has been shown in [20] that the low-complexity subcarrier independent AF MIMO-OFDM systems (i.e., the system considered in (3) and (4)) only have a slight performance loss in terms of total data mean-square-error (MSE) compared to the subcarrier cooperative AF MIMO-OFDM systems. Therefore, in this paper, we focus on the more practical subcarrier independent AF MIMO-OFDM relay systems.

III. Channel Estimation Error Modeling

In practical systems, channel state information (CSI) is unknown and must be estimated. Here, we consider estimating the channels based on training sequence. Furthermore, the two frequency-selective MIMO channels between the source and relay, and that between the relay and destination are estimated independently. In our work, the source-relay channel is estimated at the relay, while the relay-destination channel is estimated at the destination. Then each channel estimation problem is a standard point-to-point MIMO-OFDM channel estimation.

For point-to-point MIMO-OFDM systems, channels can be estimated in either frequency domain or time domain. The advantage of time domain over frequency domain channel estimation is that there are much fewer parameters to be estimated [25]. Therefore, we focus on time domain channel estimation. Because the channels in the two hops are separately estimated in time domain, we will present the first hop channel estimation as an example and the same procedure can be applied to the second hop channel estimation.

From the received signal model in frequency domain given by (3), the corresponding time domain signal is

$$\begin{aligned}\mathbf{r} &= (\boldsymbol{\mathcal{F}}^{\mathrm{H}} \otimes \mathbf{I}_{M_R})\mathbf{x} \\ &= \underbrace{(\boldsymbol{\mathcal{F}}^{\mathrm{H}} \otimes \mathbf{I}_{M_R})\mathbf{H}_{sr}(\boldsymbol{\mathcal{F}} \otimes \mathbf{I}_{N_S})}_{\triangleq \boldsymbol{\mathcal{H}}_{sr}} \underbrace{(\boldsymbol{\mathcal{F}}^{\mathrm{H}} \otimes \mathbf{I}_{N_S})\mathbf{s}}_{\triangleq \mathbf{d}} + \underbrace{(\boldsymbol{\mathcal{F}}^{\mathrm{H}} \otimes \mathbf{I}_{M_R})\mathbf{n}_1}_{\triangleq \mathbf{v}}\end{aligned} \tag{6}$$

where $\boldsymbol{\mathcal{F}}$ is the discrete-Fourier-transform (DFT) matrix with dimension $K \times K$. Based on the properties of DFT matrix, it is proved in Appendix A that (6) can be rewritten as

$$\mathbf{r} = \left(\mathbf{D}^{\mathrm{T}} \otimes \mathbf{I}_{M_R}\right) \underbrace{\mathrm{vec}([\boldsymbol{\mathcal{H}}_{sr}^{(0)} \ \cdots \ \boldsymbol{\mathcal{H}}_{sr}^{(L_1-1)}])}_{\triangleq \boldsymbol{\xi}_{sr}} + \mathbf{v}, \tag{7}$$

where the matrices $\boldsymbol{\mathcal{H}}_{sr}^{(\ell)}$ are defined as

$$\boldsymbol{\mathcal{H}}_{sr}^{(\ell)} = \sum_{k=0}^{K-1} \mathbf{H}_{sr,k} e^{j\frac{2\pi}{K}k\ell}, \qquad \ell = 0, 1, \cdots L_1 - 1. \tag{8}$$

It is obvious that $\boldsymbol{\mathcal{H}}_{sr}^{(\ell)}$ is the ℓ^{th} tap of the multi-path MIMO channel between the source and relay in the time domain and L_1 is the length of the multi-path channel. The data matrix $\mathbf{D}$ is a block circular matrix as

$$\mathbf{D} \triangleq \begin{bmatrix} \mathbf{d}_0 & \mathbf{d}_1 & \cdots & \cdots & \cdots & \mathbf{d}_{K-1} \\ \mathbf{d}_{K-1} & \mathbf{d}_0 & \ddots & \ddots & \vdots & \mathbf{d}_{K-2} \\ \vdots & \cdots & \ddots & \ddots & \vdots & \vdots \\ \mathbf{d}_{K-L_1+1} & \mathbf{d}_{K-L_1+2} & \cdots & \cdots & \cdots & \mathbf{d}_{K-L_1} \end{bmatrix}, \tag{9}$$

where the element $\mathbf{d}_i$ is expressed as

$$\mathbf{d}_i = \frac{1}{\sqrt{K}} \sum_{k=0}^{K-1} \mathbf{s}_k e^{j\frac{2\pi}{K}k(i)}, \quad i = 0, \cdots, K-1. \tag{10}$$

Based on the signal model in (7), the linear minimum-mean-square-error (LMMSE) channel estimate is given by [25]

$$\hat{\boldsymbol{\xi}}_{sr} = (\sigma_{n_1}^{-2}(\mathbf{D}^{\mathrm{T}} \otimes \mathbf{I}_{M_R})^{\mathrm{H}}(\mathbf{D}^{\mathrm{T}} \otimes \mathbf{I}_{M_R}) + \mathbf{R}_{\text{channel}}^{-1})^{-1}\sigma_{n_1}^{-2}(\mathbf{D}^{\mathrm{T}} \otimes \mathbf{I}_{M_R})^{\mathrm{H}}\mathbf{y}, \tag{11}$$

with the corresponding MSE

$$\mathbb{E}\{(\boldsymbol{\xi}_{sr} - \hat{\boldsymbol{\xi}}_{sr})(\boldsymbol{\xi}_{sr} - \hat{\boldsymbol{\xi}}_{sr})^{\mathrm{H}}\} = (\mathbf{R}_{\text{channel}}^{-1} + \sigma_{n_1}^{-2}(\mathbf{D}^*\mathbf{D}^{\mathrm{T}}) \otimes \mathbf{I}_{M_R})^{-1}, \tag{12}$$

where $\mathbf{R}_{\text{channel}} = \mathbb{E}\{\boldsymbol{\xi}_{sr}\boldsymbol{\xi}_{sr}{}^{\mathrm{H}}\}$ is the prior information for channel covariance matrix. For uncorrelated channel taps, $\mathbf{R}_{\text{channel}} = \boldsymbol{\Lambda}_{\text{channel}} \otimes \mathbf{I}_{M_R N_S}$ and $\boldsymbol{\Lambda}_{\text{channel}} = \mathrm{diag}[\sigma_{h_0}, \sigma_{h_1}, \cdots, \sigma_{h_{L-1}}]$, where σ_{h_l} is the variance of the l^{th} channel tap [24].

On the other hand, the channel in frequency domain and time domain has the following relationship[1]

$$\mathrm{vec}([\mathbf{H}_{sr,0} \ \cdots \ \mathbf{H}_{sr,K-1}]) = \sqrt{K}(\boldsymbol{\mathcal{F}}_{L_1} \otimes \mathbf{I}_{M_R N_S})\boldsymbol{\xi}_{sr}, \tag{13}$$

[1]This relationship holds for both perfect CSI and estimated CSI.

where $\boldsymbol{\mathcal{F}}_{L_1}$ is the first L_1 columns of $\boldsymbol{\mathcal{F}}$. If the frequency domain channel estimate $\hat{\mathbf{H}}_{sr,k}$ is computed according to (13), we have

$$\begin{aligned}&\mathbb{E}\{\mathrm{vec}([\Delta\mathbf{H}_{sr,0} \ \cdots \ \Delta\mathbf{H}_{sr,K-1}])\mathrm{vec}^{\mathrm{H}}([\Delta\mathbf{H}_{sr,0} \ \cdots \ \Delta\mathbf{H}_{sr,K-1}])\}\\ &= (\boldsymbol{\mathcal{F}}_{L_1}\otimes\mathbf{I}_{M_R N_S})\underbrace{(\boldsymbol{\Lambda}_{\text{channel}}^{-1}\otimes\mathbf{I}_{N_S}+\sigma_{n_1}^{-2}(\mathbf{D}^*\mathbf{D}^{\mathrm{T}}))^{-1}}_{\triangleq\boldsymbol{\Phi}^{sr}}\otimes\mathbf{I}_{M_R}(\boldsymbol{\mathcal{F}}_{L_1}\otimes\mathbf{I}_{M_R N_S})^{\mathrm{H}}K,\end{aligned} \tag{14}$$

where $\Delta\mathbf{H}_{sr,k}=\mathbf{H}_{sr,k}-\hat{\mathbf{H}}_{sr,k}$.

In case there is no prior information on $\mathbf{R}_{\text{channel}}$, we can assign uninformative prior to $\boldsymbol{\xi}_{sr}$, that is, $\sigma_{h_0},\sigma_{h_1},\cdots,\sigma_{h_{L-1}}$ approach infinity [26]. In this case, $\mathbf{R}_{\text{channel}}^{-1}\to\mathbf{0}$, and then the channel estimator (11) and estimation MSE (12) reduce to that of maximum likelihood (ML) estimation [25, P.179].

Taking the $M_R N_S\times M_R N_S$ block diagonal elements from (14) gives

$$\mathbb{E}\{\mathrm{vec}(\Delta\mathbf{H}_{sr,k})\mathrm{vec}^{\mathrm{H}}(\Delta\mathbf{H}_{sr,k})\}=\left(\sum_{\ell_2=0}^{L_1-1}\sum_{\ell_1=0}^{L_1-1}(e^{-j\frac{2\pi}{K}k(\ell_1-\ell_2)}\boldsymbol{\Phi}_{\ell_1,\ell_2}^{sr})\right)\otimes\mathbf{I}_{M_R}. \tag{15}$$

where $\boldsymbol{\Phi}_{\ell_1,\ell_2}^{sr}$ is the $N_S\times N_S$ matrix taken from the following partition of $\boldsymbol{\Phi}^{sr}$

$$\boldsymbol{\Phi}^{sr}=\begin{bmatrix}\boldsymbol{\Phi}_{0,0}^{sr} & \boldsymbol{\Phi}_{0,1}^{sr} & \cdots & \boldsymbol{\Phi}_{0,L_1-1}^{sr}\\ \vdots & \cdots & \ddots & \vdots\\ \boldsymbol{\Phi}_{L_1-1,0}^{sr} & \boldsymbol{\Phi}_{L_1-1,1}^{sr} & \cdots & \boldsymbol{\Phi}_{L_1-1,L_1-1}^{sr}\end{bmatrix}. \tag{16}$$

Furthermore, based on (15), for an arbitrary square matrix $\mathbf{R}$, it is proved in Appendix B that

$$\mathbb{E}\{\Delta\mathbf{H}_{sr,k}\mathbf{R}\Delta\mathbf{H}_{sr,k}^{\mathrm{H}}\}=\mathrm{Tr}\left(\mathbf{R}\sum_{\ell_2=0}^{L_1-1}\sum_{\ell_1=0}^{L_1-1}\left(e^{-j\frac{2\pi}{K}k(\ell_1-\ell_2)}(\boldsymbol{\Phi}_{\ell_1,\ell_2}^{sr})^{\mathrm{T}}\right)\right)\mathbf{I}_{M_R}. \tag{17}$$

A similar result holds for the second hop. In particular, denoting the relationship between the true value and estimate of the second hop channel as

$$\mathbf{H}_{rd,k}=\hat{\mathbf{H}}_{rd,k}+\Delta\mathbf{H}_{rd,k},\quad k=0,\cdots,K-1, \tag{18}$$

we have the following property

$$\mathbb{E}\{\Delta\mathbf{H}_{rd,k}\mathbf{R}\Delta\mathbf{H}_{rd,k}^{\mathrm{H}}\}=\mathrm{Tr}\left(\mathbf{R}\sum_{\ell_1=0}^{L_2-1}\sum_{\ell_2=0}^{L_2-1}\left(e^{-j\frac{2\pi}{K}k(\ell_1-\ell_2)}(\boldsymbol{\Phi}_{\ell_1,\ell_2}^{rd})^{\mathrm{T}}\right)\right)\mathbf{I}_{M_D}, \tag{19}$$

where L_2 is the length of the second hop channel in time domain. Furthermore, as the two channels are estimated independently, $\Delta\mathbf{H}_{sr,k}$ and $\Delta\mathbf{H}_{rd,k}$ are independent.

IV. Transceiver Design Problem Formulation

At the destination, a linear equalizer $\mathbf{G}_k$ is adopted for each subcarrier to detect the transmitted data $\mathbf{s}_k$ (see Fig. 1). The problem is how to design the linear precoder matrix $\mathbf{F}_k$ at the relay and the linear equalizer $\mathbf{G}_k$ at the destination to minimize the MSE of the received data at the destination:

$$\mathrm{MSE}_k(\mathbf{F}_k, \mathbf{G}_k) = \mathbb{E}\{\mathrm{Tr}\left((\mathbf{G}_k\mathbf{y}_k - \mathbf{s}_k)(\mathbf{G}_k\mathbf{y}_k - \mathbf{s}_k)^{\mathrm{H}}\right)\}, \tag{20}$$

where the expectation is taken with respect to $\mathbf{s}_k$, $\Delta\mathbf{H}_{sr,k}$, $\Delta\mathbf{H}_{rd,k}$, $\mathbf{n}_{1,k}$ and $\mathbf{n}_{2,k}$. Since $\mathbf{s}_k$, $\mathbf{n}_{1,k}$ and $\mathbf{n}_{2,k}$ are independent, the MSE expression (20) can be written as

$$\begin{aligned}
&\mathrm{MSE}_k(\mathbf{F}_k, \mathbf{G}_k)\\
&= \mathbb{E}\{\|(\mathbf{G}_k\mathbf{H}_{rd,k}\mathbf{F}_k\mathbf{H}_{sr,k} - \mathbf{I}_{N_S})\mathbf{s}_k + \mathbf{G}_k\mathbf{H}_{rd,k}\mathbf{F}_k\mathbf{n}_{1,k} + \mathbf{G}_k\mathbf{n}_{2,k}\|^2\}\\
&= \mathbb{E}_{\Delta\mathbf{H}_{sr,k},\Delta\mathbf{H}_{rd,k}}\{\mathrm{Tr}((\mathbf{G}_k\mathbf{H}_{rd,k}\mathbf{F}_k\mathbf{H}_{sr,k} - \mathbf{I}_{N_S})\mathbf{R}_{\mathbf{s}_k}(\mathbf{G}_k\mathbf{H}_{rd,k}\mathbf{F}_k\mathbf{H}_{sr,k} - \mathbf{I}_{N_S})^{\mathrm{H}})\}\\
&\quad + \mathbb{E}_{\Delta\mathbf{H}_{rd,k}}\{\mathrm{Tr}\left((\mathbf{G}_k\mathbf{H}_{rd,k}\mathbf{F}_k)\mathbf{R}_{n_{1,k}}(\mathbf{G}_k\mathbf{H}_{rd,k}\mathbf{F}_k)^{\mathrm{H}}\right)\} + \mathrm{Tr}(\mathbf{G}_k\mathbf{R}_{n_{2,k}}\mathbf{G}_k^{\mathrm{H}})\\
&= \mathbb{E}_{\Delta\mathbf{H}_{sr,k},\Delta\mathbf{H}_{rd,k}}\{\mathrm{Tr}\left((\mathbf{G}_k\mathbf{H}_{rd,k}\mathbf{F}_k\mathbf{H}_{sr,k})\mathbf{R}_{\mathbf{s}_k}(\mathbf{G}_k\mathbf{H}_{rd,k}\mathbf{F}_k\mathbf{H}_{sr,k})^{\mathrm{H}}\right)\}\\
&\quad + \mathrm{Tr}\left(\mathbf{G}_k\mathbb{E}_{\Delta\mathbf{H}_{rd,k}}\{\mathbf{H}_{rd,k}\mathbf{F}_k\mathbf{R}_{n_{1,k}}\mathbf{F}_k^{\mathrm{H}}\mathbf{H}_{rd,k}^{\mathrm{H}}\}\mathbf{G}_k^{\mathrm{H}}\right)\\
&\quad - \mathrm{Tr}\left(\mathbf{R}_{s,k}(\mathbf{G}_k\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\hat{\mathbf{H}}_{sr,k})^{\mathrm{H}}\right) - \mathrm{Tr}\left(\mathbf{G}_k\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\hat{\mathbf{H}}_{sr,k}\mathbf{R}_{s,k}\right)\\
&\quad + \mathrm{Tr}(\mathbf{R}_{\mathbf{s}_k}) + \mathrm{Tr}(\mathbf{G}_k\mathbf{R}_{n_{2,k}}\mathbf{G}_k^{\mathrm{H}}).
\end{aligned} \tag{21}$$

Because $\Delta\mathbf{H}_{sr,k}$ and $\Delta\mathbf{H}_{rd,k}$ are independent, the first term of $\mathrm{MSE_k}$ is

$$\begin{aligned}
&\mathbb{E}_{\Delta\mathbf{H}_{sr,k},\Delta\mathbf{H}_{rd,k}}\{\mathrm{Tr}\left((\mathbf{G}_k\mathbf{H}_{rd,k}\mathbf{F}_k\mathbf{H}_{sr,k})\mathbf{R}_{\mathbf{s}_k}(\mathbf{G}_k\mathbf{H}_{rd,k}\mathbf{F}_k\mathbf{H}_{sr,k})^{\mathrm{H}}\right)\}\\
&= \mathrm{Tr}\left(\mathbf{G}_k\mathbb{E}_{\Delta\mathbf{H}_{rd,k}}\left\{\mathbf{H}_{rd,k}\mathbf{F}_k\mathbb{E}_{\Delta\mathbf{H}_{sr,k}}\{\mathbf{H}_{sr,k}\mathbf{R}_{\mathbf{s}_k}\mathbf{H}_{sr,k}^{\mathrm{H}}\}\mathbf{F}_k^{\mathrm{H}}\mathbf{H}_{rd,k}^{\mathrm{H}}\right\}\mathbf{G}_k^{\mathrm{H}}\right).
\end{aligned} \tag{22}$$

For the inner expectation, the following equation holds

$$\begin{aligned}
\mathbb{E}_{\Delta\mathbf{H}_{sr,k}}\{\mathbf{H}_{sr,k}\mathbf{R}_{\mathbf{s}_k}\mathbf{H}_{sr,k}^{\mathrm{H}}\} &= \mathbb{E}_{\Delta\mathbf{H}_{sr,k}}\{(\hat{\mathbf{H}}_{sr,k} + \Delta\mathbf{H}_{sr,k})\mathbf{R}_{\mathbf{s}_k}(\hat{\mathbf{H}}_{sr,k} + \Delta\mathbf{H}_{sr,k})^{\mathrm{H}}\}\\
&= \mathrm{Tr}(\mathbf{R}_{\mathbf{s}_k}\mathbf{\Psi}_{sr,k})\mathbf{I}_{M_R} + \hat{\mathbf{H}}_{sr,k}\mathbf{R}_{\mathbf{s}_k}\hat{\mathbf{H}}_{sr,k}^{\mathrm{H}} \triangleq \mathbf{\Pi}_k,
\end{aligned} \tag{23}$$

where based on (17) the matrix $\mathbf{\Psi}_{sr,k}$ is defined as

$$\mathbf{\Psi}_{sr,k} = \sum_{\ell_1=0}^{L_1-1}\sum_{\ell_2=0}^{L_1-1}\left(e^{-j\frac{2\pi}{K}k(\ell_1-\ell_2)}(\mathbf{\Phi}_{\ell_1,\ell_2}^{sr})^{\mathrm{T}}\right). \tag{24}$$

Applying (23) and the corresponding result for $\Delta\mathbf{H}_{rd,k}$ to (22), the first term of MSE_k becomes

$$\begin{aligned}&\mathrm{Tr}\left(\mathbf{G}_k\mathbb{E}_{\Delta\mathbf{H}_{rd,k}}\left\{\mathbf{H}_{rd,k}\mathbf{F}_k\mathbb{E}_{\Delta\mathbf{H}_{sr,k}}\{\mathbf{H}_{sr,k}\mathbf{R}_{\mathbf{s}_k}\mathbf{H}_{sr,k}^{\mathrm{H}}\}\mathbf{F}_k^{\mathrm{H}}\mathbf{H}_{rd,k}^{\mathrm{H}}\right\}\mathbf{G}_k^{\mathrm{H}}\right)\\&=\mathrm{Tr}(\mathbf{G}_k(\mathrm{Tr}(\mathbf{F}_k\mathbf{\Pi}_k\mathbf{F}_k^{\mathrm{H}}\mathbf{\Psi}_{rd,k})\mathbf{I}_{M_D}+\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\mathbf{\Pi}_k\mathbf{F}_k^{\mathrm{H}}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}})\mathbf{G}_k^{\mathrm{H}}),\end{aligned} \tag{25}$$

where the matrix $\mathbf{\Psi}_{rd,k}$ is defined as

$$\mathbf{\Psi}_{rd,k}=\sum_{\ell_1=0}^{L_2-1}\sum_{\ell_2=0}^{L_2-1}\left(e^{-j\frac{2\pi}{K}k(\ell_1-\ell_2)}(\mathbf{\Phi}_{\ell_1,\ell_2}^{rd})^{\mathrm{T}}\right). \tag{26}$$

Similarly, the second term of MSE_k in (21) can be simplified as

$$\begin{aligned}&\mathrm{Tr}\left(\mathbf{G}_k\mathbb{E}_{\Delta\mathbf{H}_{rd,k}}\{\mathbf{H}_{rd,k}\mathbf{F}_k\mathbf{R}_{n_1,k}\mathbf{F}_k^{\mathrm{H}}\mathbf{H}_{rd,k}^{\mathrm{H}}\}\mathbf{G}_k^{\mathrm{H}}\right)\\&=\mathrm{Tr}(\mathbf{G}_k\left(\mathrm{Tr}(\mathbf{F}_k\mathbf{R}_{n_1,k}\mathbf{F}_k^{\mathrm{H}}\mathbf{\Psi}_{rd,k})\mathbf{I}_{M_D}+\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\mathbf{R}_{n_1,k}\mathbf{F}_k^{\mathrm{H}}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}}\right)\mathbf{G}_k^{\mathrm{H}}).\end{aligned} \tag{27}$$

Based on (25) and (27), the MSE_k (21) equals to

$$\begin{aligned}\mathrm{MSE}_k(\mathbf{F}_k,\mathbf{G}_k)=&\ \mathrm{Tr}\left(\mathbf{G}_k(\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\mathbf{F}_k^{\mathrm{H}}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}}+\mathbf{K}_k)\mathbf{G}_k^{\mathrm{H}}\right)-\mathrm{Tr}\left(\mathbf{R}_{\mathbf{s}_k}\hat{\mathbf{H}}_{sr,k}^{\mathrm{H}}\mathbf{F}_k^{\mathrm{H}}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}}\right)\\&-\mathrm{Tr}\left(\mathbf{G}_k\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\hat{\mathbf{H}}_{sr,k}\mathbf{R}_{\mathbf{s}_k}\right)+\mathrm{Tr}(\mathbf{R}_{\mathbf{s}_k})\end{aligned} \tag{28}$$

where

$$\mathbf{R}_{\mathbf{x}_k}=\mathbf{\Pi}_k+\sigma_{n_1}^2\mathbf{I}_{M_R} \tag{29}$$

$$\begin{aligned}\mathbf{K}_k&=(\mathrm{Tr}(\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\mathbf{F}_k^{\mathrm{H}}\mathbf{\Psi}_{rd,k})+\sigma_{n_2}^2)\mathbf{I}_{M_D}\\&\triangleq\eta_k\mathbf{I}_{M_D}.\end{aligned} \tag{30}$$

Notice that the matrix $\mathbf{R}_{\mathbf{x}_k}$ is the correlation matrix of the receive signal $\mathbf{x}_k$ on the k^{th} subcarrier at the relay.

Subject to the transmit power constraint at the relay, the joint design of precoder at the relay and equalizer at the destination that minimizes the total MSE of the output data at the destination can be formulated as the following optimization problem

$$\begin{aligned}\min_{\mathbf{F}_k,\mathbf{G}_k}\quad&\sum_k\mathrm{MSE}_k(\mathbf{F}_k,\mathbf{G}_k)\\\mathrm{s.t.}\quad&\sum_k\mathrm{Tr}(\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\mathbf{F}_k^{\mathrm{H}})\le P_r.\end{aligned} \tag{31}$$

Remark 1: In this paper, the relay estimates the source-relay channel and the destination estimates the relay-destination channel. The precoder $\mathbf{F}_k$ and $\mathbf{G}_k$ can be designed at the relay or at the destination. Sharing channel estimation between the relay and the destination is unavoidable. However, when channel is varying slowly, and the channel estimation feedback occurs infrequently, the errors in feedback can be negligible.

V. Proposed Closed-Form Solution

In this section, we will derive a closed-form solution for the optimization problem (31). In order to facilitate the analysis, the optimization problem (31) is rewritten as

$$\begin{aligned} \min_{\mathbf{F}_k,\mathbf{G}_k,P_{r,k}} \quad & \sum_k \mathrm{MSE}_k(\mathbf{F}_k,\mathbf{G}_k) \\ \text{s.t.} \quad & \mathrm{Tr}(\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\mathbf{F}_k^{\mathrm{H}}) \leq P_{r,k}, \qquad k=0,\cdots,K-1 \\ & \sum_k P_{r,k} \leq P_r, \end{aligned} \tag{32}$$

with the physical meaning of $P_{r,k}$ being the maximum allocated power over the k^{th} subcarrier.

The Lagrangian function of the optimization problem (32) is

$$\mathcal{L}(\mathbf{F}_k,\mathbf{G}_k,P_{r,k}) = \sum_k \mathrm{MSE}_k(\mathbf{F}_k,\mathbf{G}_k) + \sum_k \gamma_k(\mathrm{Tr}(\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\mathbf{F}_k^{\mathrm{H}}) - P_{r,k}) + \gamma(\sum_k P_{r,k} - P_r) \tag{33}$$

where the positive scalars γ_k and γ are the Lagrange multipliers. Differentiating (33) with respect to $\mathbf{F}_k$, $\mathbf{G}_k$ and $P_{r,k}$, and setting the corresponding results to zero, the Karush-Kuhn-Tucker (KKT) conditions of the optimization problem (32) are given by [27]

$$\mathbf{G}_k(\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\mathbf{F}_k^{\mathrm{H}}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}} + \mathbf{K}_k) = \mathbf{R}_{\mathbf{s}_k}(\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\hat{\mathbf{H}}_{sr,k})^{\mathrm{H}}, \tag{34a}$$

$$\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}}\mathbf{G}_k\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k} + (\mathrm{Tr}(\mathbf{G}_k\mathbf{G}_k^{\mathrm{H}})\mathbf{\Psi}_{rd,k} + \gamma_k)\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k} = \left(\hat{\mathbf{H}}_{sr,k}\mathbf{R}_{\mathbf{s}_k}\mathbf{G}_k\hat{\mathbf{H}}_{rd,k}\right)^{\mathrm{H}}, \tag{34b}$$

$$\gamma_k(\mathrm{Tr}(\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\mathbf{F}_k^{\mathrm{H}}) - P_{r,k}) = 0, \tag{34c}$$

$$\gamma_k \geq 0, \qquad k=0,\cdots,K-1, \tag{34d}$$

$$\gamma(\sum_k P_{r,k} - P_r) = 0, \tag{34e}$$

$$\gamma_0 = \gamma_1 = \cdots = \gamma_{K-1} = \gamma, \tag{34f}$$

$$\mathrm{Tr}(\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\mathbf{F}_k^{\mathrm{H}}) \leq P_{r,k}, \tag{34g}$$

$$\sum_k P_{r,k} \leq P_r. \tag{34h}$$

It is obvious that the objective function and constraints of (32) are continuously differentiable. Furthermore, it is easy to see that solutions of the optimization problem (32) satisfy the regularity condition, i.e., Abadie constraint qualification (ACQ), because linear independence constraint qualification (LICQ) can be proved [28]. Based on these facts, the KKT conditions are the necessary conditions.[2] From KKT conditions, we can have the following two useful properties which can help us to find the optimal solution.

Property 1: It is proved in Appendix C that for any $\mathbf{F}_k$ satisfying the KKT conditions (34a)-(34e), the power constraints (34g) and (34h) must occur on the boundaries

$$\mathrm{Tr}(\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\mathbf{F}_k^{\mathrm{H}}) = P_{r,k}, \tag{35}$$

$$\sum_k P_{r,k} = P_r. \tag{36}$$

Furthermore, the corresponding $\mathbf{G}_k$ satisfies

$$\mathrm{Tr}(\mathbf{G}_k\mathbf{G}_k^{\mathrm{H}}) = \gamma_k P_{r,k}/\sigma_{n_2}^2. \tag{37}$$

Property 2: Define the matrices $\mathbf{U}_{\mathbf{T}_k}$, $\mathbf{V}_{\mathbf{T}_k}$, $\mathbf{\Lambda}_{\mathbf{T}_k}$, $\mathbf{U}_{\mathbf{\Theta}_k}$, and $\mathbf{\Lambda}_{\mathbf{\Theta}_k}$ based on singular value decomposition (SVD) as

$$(P_{r,k}\mathbf{\Psi}_{rd,k} + \sigma_{n_2}^2\mathbf{I}_{N_R})^{-\frac{\mathrm{H}}{2}}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}}\underbrace{\hat{\mathbf{H}}_{rd,k}(P_{r,k}\mathbf{\Psi}_{rd,k} + \sigma_{n_2}^2\mathbf{I}_{N_R})^{-\frac{1}{2}}}_{\triangleq \mathbf{\Theta}_k} = \mathbf{U}_{\mathbf{\Theta}_k}\mathbf{\Lambda}_{\mathbf{\Theta}_k}\mathbf{U}_{\mathbf{\Theta}_k}^{\mathrm{H}}, \tag{38}$$

$$\mathbf{R}_{\mathbf{x},k}^{-\frac{1}{2}}\hat{\mathbf{H}}_{sr,k}\mathbf{R}_{\mathbf{s}_k} = \mathbf{U}_{\mathbf{T}_k}\mathbf{\Lambda}_{\mathbf{T}_k}\mathbf{V}_{\mathbf{T}_k}^{\mathrm{H}}, \tag{39}$$

with elements of the diagonal matrix $\mathbf{\Lambda}_{\mathbf{T}_k}$ and $\mathbf{\Lambda}_{\mathbf{\Theta}_k}$ arranged in decreasing order. Then with KKT conditions (34a) and (34b), it is proved in Appendix D that the optimal precoder $\mathbf{F}_k$ and equalizer $\mathbf{G}_k$ for the optimization problem (32) are in the forms of

$$\mathbf{F}_k = (P_{r,k}\mathbf{\Psi}_{rd,k} + \sigma_{n_2}^2\mathbf{I}_{N_R})^{-\frac{1}{2}}\mathbf{U}_{\mathbf{\Theta}_k,q_k}\boldsymbol{A}_{\mathbf{F}_k}\mathbf{U}_{\mathbf{T}_k,p_k}^{\mathrm{H}}\mathbf{R}_{\mathbf{x}_k}^{-\frac{1}{2}}, \tag{40}$$

$$\mathbf{G}_k = \mathbf{V}_{\mathbf{T}_k,p_k}\boldsymbol{A}_{\mathbf{G}_k}\mathbf{U}_{\mathbf{\Theta}_k,q_k}^{\mathrm{H}}(P_{r,k}\mathbf{\Psi}_{rd,k} + \sigma_{n_2}^2\mathbf{I}_{N_R})^{-\frac{\mathrm{H}}{2}}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}}, \tag{41}$$

[2]Notice that the solution $\mathbf{F}_0 = \cdots = \mathbf{F}_{K-1} = \mathbf{0}$ and $\mathbf{G}_0 = \cdots = \mathbf{G}_{K-1} = \mathbf{0}$ also satisfies the KKT conditions, but this solution is meaningless as no signal can be transmitted [14].

where $\boldsymbol{A}_{\mathbf{F}_k}$ and $\boldsymbol{A}_{\mathbf{G}_k}$ are to be determined. The matrix $\mathbf{U}_{\mathbf{T}_k,p_k}$ and $\mathbf{V}_{\mathbf{T}_k,p_k}$ are the first p_k columns of $\mathbf{U}_{\mathbf{T}_k}$ and $\mathbf{V}_{\mathbf{T}_k}$, respectively, and $p_k = \mathrm{Rank}(\boldsymbol{\Lambda}_{\mathbf{T}_k})$. Similarly, $\mathbf{U}_{\boldsymbol{\Theta}_k,q_k}$ is the first q_k columns of $\mathbf{U}_{\boldsymbol{\Theta}_k}$, and $q_k = \mathrm{Rank}(\boldsymbol{\Lambda}_{\boldsymbol{\Theta}_k})$.

Right multiplying both sides of (34a) with $\mathbf{G}_k^{\mathrm{H}}$ and left multiplying both sides of (34b) with $\mathbf{F}_k^{\mathrm{H}}$, and making use of (40) and (41), the first two KKT conditions become

$$\boldsymbol{A}_{\mathbf{G}_k}\bar{\boldsymbol{\Lambda}}_{\boldsymbol{\Theta}_k}\boldsymbol{A}_{\mathbf{F}_k}\boldsymbol{A}_{\mathbf{F}_k}^{\mathrm{H}}\bar{\boldsymbol{\Lambda}}_{\boldsymbol{\Theta}_k}\boldsymbol{A}_{\mathbf{G}_k}^{\mathrm{H}} + \eta_k\boldsymbol{A}_{\mathbf{G}_k}\bar{\boldsymbol{\Lambda}}_{\boldsymbol{\Theta}_k}\boldsymbol{A}_{\mathbf{G}_k}^{\mathrm{H}} = (\boldsymbol{A}_{\mathbf{G}_k}\bar{\boldsymbol{\Lambda}}_{\boldsymbol{\Theta}_k}\boldsymbol{A}_{\mathbf{F}_k}\bar{\boldsymbol{\Lambda}}_{\mathbf{T}_k})^{\mathrm{H}}, \tag{42}$$

$$\boldsymbol{A}_{\mathbf{F}_k}^{\mathrm{H}}\bar{\boldsymbol{\Lambda}}_{\boldsymbol{\Theta}_k}\boldsymbol{A}_{\mathbf{G}_k}^{\mathrm{H}}\boldsymbol{A}_{\mathbf{G}_k}\bar{\boldsymbol{\Lambda}}_{\boldsymbol{\Theta}_k}\boldsymbol{A}_{\mathbf{F}_k} + \frac{\gamma_k}{\sigma_{n_2}^2}\boldsymbol{A}_{\mathbf{F}_k}^{\mathrm{H}}\boldsymbol{A}_{\mathbf{F}_k} = (\bar{\boldsymbol{\Lambda}}_{\mathbf{T}_k}\boldsymbol{A}_{\mathbf{G}_k}\bar{\boldsymbol{\Lambda}}_{\boldsymbol{\Theta}_k}\boldsymbol{A}_{\mathbf{F}_k})^{\mathrm{H}}, \tag{43}$$

where the matrix $\bar{\boldsymbol{\Lambda}}_{\boldsymbol{\Theta}_k}$ is the $q_k \times q_k$ principal submatrix of $\boldsymbol{\Lambda}_{\boldsymbol{\Theta}_k}$. Similarly, $\bar{\boldsymbol{\Lambda}}_{\mathbf{T}_k}$ is the $p_k \times p_k$ principal submatrix of $\boldsymbol{\Lambda}_{\mathbf{T}_k}$. In this paper, we consider AF MIMO-OFDM relay systems, the matrices $\boldsymbol{A}_{\mathbf{F}_k}$ and $\boldsymbol{A}_{\mathbf{G}_k}$ can be of arbitrary dimension instead of the square matrices considered in [14] and [22]. It should be noticed that as the optimization problem (32) is not a convex problem, the KKT conditions are only necessary conditions. That is, there are many solutions that will satisfy the KKT conditions. To identify the optimal solution, we need an additional information which is presented in the following ***Property 3***.

Property 3: Putting the results of ***Property 1*** and ***Property 2*** into the optimization problem (32), based on majorization theory, it is proved in Appendix E that the optimal $\boldsymbol{A}_{\mathbf{F}_k}$ and $\boldsymbol{A}_{\mathbf{G}_k}$ have the following diagonal structure

$$\boldsymbol{A}_{\mathbf{F}_k,\mathrm{opt}} = \begin{bmatrix} \boldsymbol{\Lambda}_{\mathbf{F}_k,\mathrm{opt}} & \mathbf{0}_{N_k,p_k-N_k} \\ \mathbf{0}_{q_k-N_k,N_k} & \mathbf{0}_{q_k-N_k,p_k-N_k} \end{bmatrix}, \tag{44}$$

$$\boldsymbol{A}_{\mathbf{G}_k,\mathrm{opt}} = \begin{bmatrix} \boldsymbol{\Lambda}_{\mathbf{G}_k,\mathrm{opt}} & \mathbf{0}_{N_k,q_k-N_k} \\ \mathbf{0}_{p_k-N_k,N_k} & \mathbf{0}_{p_k-N_k,q_k-N_k} \end{bmatrix}, \tag{45}$$

where $\boldsymbol{\Lambda}_{\mathbf{F}_k,\mathrm{opt}}$ and $\boldsymbol{\Lambda}_{\mathbf{G}_k,\mathrm{opt}}$ are two $N_k \times N_k$ diagonal matrices to be determined, and $N_k = \min(p_k, q_k)$.

Substituting (44) and (45) into (42) and (43), and noticing that all matrices are diagonal,

$\mathbf{\Lambda}_{\mathbf{F}_k,\text{opt}}$ and $\mathbf{\Lambda}_{\mathbf{G}_k,\text{opt}}$ can be easily solved to be

$$\mathbf{\Lambda}_{\mathbf{F}_k,\text{opt}} = \left[\left(\sqrt{\frac{\sigma_{n_2}^2\eta_k}{\gamma_k}}\tilde{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}^{-\frac{1}{2}}\tilde{\mathbf{\Lambda}}_{\mathbf{T}_k} - \eta_k\tilde{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}^{-1}\right)^+\right]^{\frac{1}{2}}, \tag{46}$$

$$\mathbf{\Lambda}_{\mathbf{G_k},\text{opt}} = \left[\left(\sqrt{\frac{\gamma_k}{\eta_k\sigma_{n_2}^2}}\tilde{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}^{-\frac{1}{2}}\tilde{\mathbf{\Lambda}}_{\mathbf{T}_k} - \frac{\gamma_k}{\sigma_{n_2}^2}\tilde{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}^{-1}\right)^+\right]^{\frac{1}{2}}\tilde{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}^{-\frac{1}{2}}, \tag{47}$$

where the matrices $\tilde{\mathbf{\Lambda}}_{\mathbf{T}_k}$ and $\tilde{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}$ are the principal sub-matrices of $\mathbf{\Lambda}_{\mathbf{T}_k}$ and $\mathbf{\Lambda}_{\mathbf{\Theta}_k}$ with dimension $N_k \times N_k$, and $N_k = \min\{\text{rank}(\mathbf{\Lambda}_{\mathbf{\Theta}_k}), \text{rank}(\mathbf{\Lambda}_{\mathbf{T}_k})\}$. The matrices $\mathbf{U}_{\mathbf{T}_k,N_k}$, $\mathbf{V}_{\mathbf{T}_k,N_k}$ and $\mathbf{U}_{\mathbf{\Theta}_k,N_k}$ are the first N_k columns of $\mathbf{U}_{\mathbf{T}_k}$, $\mathbf{V}_{\mathbf{T}_k}$ and $\mathbf{U}_{\mathbf{\Theta}_k}$, respectively.

In the general solution (46)-(47), $P_{r,k}$, η_k and γ_k are unknown. However notice that from (35) and (37) in ***Property 1***, the optimal precoder and equalizer should simultaneously satisfy

$$\text{Tr}(\mathbf{F}_{k,\text{opt}}\mathbf{R}_{\mathbf{x}_k}\mathbf{F}_{k,\text{opt}}^{\text{H}}) = P_{r,k}, \tag{48}$$

$$\text{Tr}(\mathbf{G}_{k,\text{opt}}\mathbf{G}_{k,\text{opt}}^{\text{H}}) = \gamma_k P_{r,k}/\sigma_{n_2}^2. \tag{49}$$

Substituting (44)-(47) into (48) and (49), it can be straightforwardly shown that η_k and γ_k can be expressed as functions of $P_{r,k}$

$$\eta_k = \frac{b_{3,k}P_{r,k}}{P_{r,k}b_{1,k} + b_{1,k}b_{4,k} - b_{2,k}b_{3,k}}, \tag{50}$$

$$\gamma_k = \frac{b_{3,k}\sigma_{n_2}^2(P_{r,k}b_{1,k} + b_{1,k}b_{4,k} - b_{2,k}b_{3,k})}{(P_{r,k}+b_{4,k})^2P_{r,k}}, \tag{51}$$

where $b_{1,k}$, $b_{2,k}$, $b_{3,k}$ and $b_{4,k}$ are defined as

$$b_{1,k} \triangleq \text{Tr}(\mathbf{U}_{\mathbf{\Theta}_k,N_k}^{\text{H}}(P_{r,k}\mathbf{\Psi}_{rd,k} + \sigma_{n_2}^2\mathbf{I}_{N_R})^{-1}\mathbf{U}_{\mathbf{\Theta}_k,N_k}\tilde{\mathbf{\Lambda}}_{\mathbf{T}_k}\tilde{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}^{-\frac{1}{2}}\mathbf{\Lambda}_{\mathbf{I},k}), \tag{52a}$$

$$b_{2,k} \triangleq \text{Tr}(\mathbf{U}_{\mathbf{\Theta}_k,N_k}^{\text{H}}(P_{r,k}\mathbf{\Psi}_{rd,k} + \sigma_{n_2}^2\mathbf{I}_{N_R})^{-1}\mathbf{U}_{\mathbf{\Theta}_k,N_k}\tilde{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}^{-1}\mathbf{\Lambda}_{\mathbf{I},k}), \tag{52b}$$

$$b_{3,k} \triangleq \text{Tr}(\tilde{\mathbf{\Lambda}}_{\mathbf{T}_k}\tilde{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}^{-\frac{1}{2}}\mathbf{\Lambda}_{\mathbf{I},k}), \tag{52c}$$

$$b_{4,k} \triangleq \text{Tr}(\tilde{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}^{-1}\mathbf{\Lambda}_{\mathbf{I},k}), \tag{52d}$$

and $\mathbf{\Lambda}_{\mathbf{I},k}$ is a diagonal selection matrix with diagonal elements being 1 or 0, and serves to replace the operation '+'. Combining all the results in this section, we have the following summary.

Summary: The optimal precoder $\mathbf{F}_{k,\text{opt}}$ and equalizer $\mathbf{G}_{k,\text{opt}}$ are

$$\mathbf{F}_{k,\text{opt}} = (P_{r,k}\mathbf{\Psi}_{rd,k} + \sigma_{n_2}^2\mathbf{I}_{N_R})^{-\frac{1}{2}}\mathbf{U}_{\mathbf{\Theta}_k,N_k}\mathbf{\Lambda}_{\mathbf{F}_k,\text{opt}}\mathbf{U}^{\text{H}}_{\mathbf{T}_k,N_k}\mathbf{R}_{\mathbf{x}_k}^{-\frac{1}{2}}, \tag{53}$$

$$\mathbf{G}_{k,\text{opt}} = \mathbf{V}_{\mathbf{T}_k,N_k}\mathbf{\Lambda}_{\mathbf{G}_k,\text{opt}}\mathbf{U}^{\text{H}}_{\mathbf{\Theta}_k,N_k}(P_{r,k}\mathbf{\Psi}_{rd,k} + \sigma_{n_2}^2\mathbf{I}_{N_R})^{-\frac{\text{H}}{2}}\hat{\mathbf{H}}^{\text{H}}_{rd,k}, \tag{54}$$

where

$$\mathbf{\Lambda}_{\mathbf{F}_k,\text{opt}} = \left[\left(\sqrt{\frac{\sigma_{n_2}^2\eta_k}{\gamma_k}}\tilde{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}^{-\frac{1}{2}}\tilde{\mathbf{\Lambda}}_{\mathbf{T}_k} - \eta_k\tilde{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}^{-1}\right)^{+}\right]^{\frac{1}{2}}, \tag{55}$$

$$\mathbf{\Lambda}_{\mathbf{G_k},\text{opt}} = \left[\left(\sqrt{\frac{\gamma_k}{\eta_k\sigma_{n_2}^2}}\tilde{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}^{-\frac{1}{2}}\tilde{\mathbf{\Lambda}}_{\mathbf{T}_k} - \frac{\gamma_k}{\sigma_{n_2}^2}\tilde{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}^{-1}\right)^{+}\right]^{\frac{1}{2}}\tilde{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}^{-\frac{1}{2}}, \tag{56}$$

with η_k and γ_k given by (50)-(52).

From the above summary, it is obvious that the problem of finding optimal precoder and equalizer reduces to computing $P_{r,k}$, and it can be solved based on (51) and the following two constraints (i.e., (34f) and (36))

$$\gamma_0 = \cdots = \gamma_{K-1}, \tag{57}$$

$$\sum_k P_{r,k} = P_r. \tag{58}$$

In the following subsections, we will discuss how to compute $P_{r,k}$.

Remark 2: When both channels in the two hops are flat-fading channels, the considered system reduces to single-carrier AF MIMO relay system. It should be noticed that for single-carrier MIMO relay systems, there is no need to consider power allocation among subcarriers, and we can set $P_{r,k} = P_r$. In this case, the proposed closed-form solution is exactly the optimal solution for the robust transceiver design in flat-fading channel. Furthermore, when the CSI in the two hops are perfectly known, the derived solution reduces to the optimal solution proposed in [19].

Remark 3: Notice that when the source-relay link is noiseless and the first hop channel is an identity matrix, the closed-form solution can be simplified to the optimal robust linear MMSE transceiver for point-to-point MIMO-OFDM systems [24]. Moreover, if single carrier transmission is employed, the closed-form solution further reduces to the optimal point-to-point MIMO robust LMMSE transceiver [22].

Remark 4: The complexity of the proposed algorithm is dominated matrix decomposition, matrix multiplication and matrix inversion. The complexity of each of these matrix operations is known to be $O(n^3)$, where n is the matrix dimension [29]. So the complexity of our proposed algorithm is $O(m^3)$, where $m = \max\{M_D, N_R, N_D, N_S\}$.

A. Uncorrelated Channel Estimation Error

When the channel estimation errors are uncorrelated (for example, by using training sequences that are white in both time and space dimensions), the following condition must be satisfied [10], [30]–[32]

$$\mathbf{D}\mathbf{D}^{\mathrm{H}} \propto \mathbf{I}_{N_S L_1}. \tag{59}$$

Then according to (14), we have $\mathbf{\Psi}_{sr,k} = \sum_{\ell_1} \mathbf{\Phi}^{sr}_{\ell_1,\ell_1}/K \propto \mathbf{I}_{N_S}$. Similarly, for the second hop, we also have

$$\mathbf{\Psi}_{rd,k} \propto \mathbf{I}_{N_R} \triangleq \delta_{rd,k}\mathbf{I}_{N_R}, \tag{60}$$

where the specific form of $\delta_{rd,k}$ can be easily derived based on (26).

Putting (60) into the left hand side of (38), the expression becomes

$$\begin{aligned}&(P_{r,k}\mathbf{\Psi}_{rd,k} + \sigma^2_{n_2}\mathbf{I}_{N_R})^{-\frac{\mathrm{H}}{2}}\hat{\mathbf{H}}^{\mathrm{H}}_{rd,k}\hat{\mathbf{H}}_{rd,k}(P_{r,k}\mathbf{\Psi}_{rd,k} + \sigma^2_{n_2}\mathbf{I}_{N_R})^{-\frac{1}{2}}\\ &= \frac{1}{P_{r,k}\delta_{rd,k} + \sigma^2_{n_2}}\hat{\mathbf{H}}^{\mathrm{H}}_{rd,k}\hat{\mathbf{H}}_{rd,k}.\end{aligned} \tag{61}$$

Applying eigen-decomposition $\hat{\mathbf{H}}^{\mathrm{H}}_{rd,k}\hat{\mathbf{H}}_{rd,k} = \mathbf{U}_{\mathbf{H}_k}\mathbf{\Lambda}_{\mathbf{H}_k}\mathbf{U}^{\mathrm{H}}_{\mathbf{H}_k}$ and comparing with the right hand side of (38), we have

$$\mathbf{U}_{\mathbf{\Theta}_k} = \mathbf{U}_{\mathbf{H}_k}, \quad \mathbf{\Lambda}_{\mathbf{\Theta}_k} = \frac{1}{(P_{r,k}\delta_{rd,k} + \sigma^2_{n_2})}\mathbf{\Lambda}_{\mathbf{H}_k}. \tag{62}$$

Substituting (62) into (51), γ_k reduces to

$$\gamma_k = \frac{\sigma^2_{n_2}\left(\mathrm{Tr}(\tilde{\mathbf{\Lambda}}_{\mathbf{T}_k}\tilde{\mathbf{\Lambda}}^{-\frac{1}{2}}_{\mathbf{H}_k}\mathbf{\Lambda}_{\mathbf{I},k})\right)^2}{\left(P_{r,k}\left(1 + \delta_{rd,k}\mathrm{Tr}(\tilde{\mathbf{\Lambda}}^{-1}_{\mathbf{H}_k}\mathbf{\Lambda}_{\mathbf{I},k})\right) + \sigma^2_{n_2}\mathrm{Tr}(\tilde{\mathbf{\Lambda}}^{-1}_{\mathbf{H}_k}\mathbf{\Lambda}_{\mathbf{I},k})\right)^2}, \tag{63}$$

where $\tilde{\mathbf{\Lambda}}_{\mathbf{H}_k}$ is the $N_k \times N_k$ principal submatrix of $\mathbf{\Lambda}_{\mathbf{H}_k}$.

With (63) and the facts that $\sum_k P_{r,k} = P_r$ and $\gamma_0 = \cdots = \gamma_{K-1}$, $P_{r,k}$ can be straightforwardly computed to be

$$P_{r,k} = \sqrt{\frac{\sigma_{n_2}^2}{\gamma}} \frac{\mathrm{Tr}(\tilde{\mathbf{\Lambda}}_{\mathbf{T}_k} \tilde{\mathbf{\Lambda}}_{\mathbf{H}_k}^{-\frac{1}{2}} \mathbf{\Lambda}_{\mathbf{I},k})}{1 + \delta_{rd,k} \mathrm{Tr}(\tilde{\mathbf{\Lambda}}_{\mathbf{H}_k}^{-1} \mathbf{\Lambda}_{\mathbf{I},k})} - \frac{\sigma_{n_2}^2 \mathrm{Tr}(\tilde{\mathbf{\Lambda}}_{\mathbf{H}_k}^{-1} \mathbf{\Lambda}_{\mathbf{I},k})}{1 + \delta_{rd,k} \mathrm{Tr}(\tilde{\mathbf{\Lambda}}_{\mathbf{H}_k}^{-1} \mathbf{\Lambda}_{\mathbf{I},k})}, \qquad k = 0, \cdots K-1, \tag{64}$$

where γ equals to

$$\gamma = \sigma_{n_2}^2 \left(\sum_k \frac{\mathrm{Tr}(\tilde{\mathbf{\Lambda}}_{\mathbf{T}_k} \tilde{\mathbf{\Lambda}}_{\mathbf{H}_k}^{-\frac{1}{2}} \mathbf{\Lambda}_{\mathbf{I},k})}{1 + \delta_{rd,k} \mathrm{Tr}(\tilde{\mathbf{\Lambda}}_{\mathbf{H}_k}^{-1} \mathbf{\Lambda}_{\mathbf{I},k})} \right)^2 \Bigg/ \left(P_r + \sum_k \frac{\sigma_{n_2}^2 \mathrm{Tr}(\tilde{\mathbf{\Lambda}}_{\mathbf{H}_k}^{-1} \mathbf{\Lambda}_{\mathbf{I},k})}{1 + \delta_{rd,k} \mathrm{Tr}(\tilde{\mathbf{\Lambda}}_{\mathbf{H}_k}^{-1} \mathbf{\Lambda}_{\mathbf{I},k})} \right)^2. \tag{65}$$

B. Correlated Channel Estimation Error

Due to limited length of training sequence, $\mathbf{D}\mathbf{D}^{\mathrm{H}} \propto \mathbf{I}$ may not be possible to achieve [30]. In this case, the channel estimation errors are correlated, and $\mathbf{\Psi}_{rd,k} \not\propto \mathbf{I}$. From (38), it can be seen that the relationship between $\mathbf{\Lambda}_{\mathbf{\Theta}_k}$ and $P_{r,k}$ cannot be expressed in a closed-form . Then the solution for $P_{r,k}$ cannot be directly obtained. However, notice that when $P_{r,k}\lambda_{\min}(\mathbf{\Psi}_{rd,k}) \gg \sigma_{n_2}^2$, where $\lambda_{\min}(\mathbf{Z})$ denotes the minimum eigenvalue of $\mathbf{Z}$, we have

$$P_{r,k}\mathbf{\Psi}_{rd,k} + \sigma_{n_2}^2 \mathbf{I}_{N_R} \approx P_{r,k}\mathbf{\Psi}_{rd,k}. \tag{66}$$

This situation occurs at high SNR in the second hop, and we term this high SNR approximation (HSA). On the other hand, when $P_{r,k}\lambda_{\min}(\mathbf{\Psi}_{rd,k}) \gg \sigma_{n_2}^2$ cannot be guaranteed, we have

$$P_{r,k}\mathbf{\Psi}_{rd,k} + \sigma_{n_2}^2 \mathbf{I}_{N_R} \approx (P_{r,k}\lambda_{\max}(\mathbf{\Psi}_{rd,k}) + \sigma_{n_2}^2)\mathbf{I}_{N_R}, \tag{67}$$

and it is termed spectral approximation (SPA). For spectral approximation, $\mathbf{\Psi}_{rd}$ is replaced by $\lambda_{\max}(\mathbf{\Psi}_{rd})\mathbf{I}$, and from the MSE formulation in (28), it is obvious that the resultant expression forms an upper-bound to the original MSE. Notice that when the training sequences are close to white sequence [34], [35], the eigenvalue spread of $\mathbf{\Psi}_{rd}$ is small, and SPA is a good approximation. In the following, computations of $P_{r,k}$ under different approximations are detailed.

1) High SNR Approximation (HSA) at the second hop: Based on HSA, the left hand side of (38) becomes

$$\begin{aligned}
&(P_{r,k}\mathbf{\Psi}_{rd,k} + \sigma_{n_2}^2 \mathbf{I}_{N_R})^{-\frac{\mathrm{H}}{2}} \hat{\mathbf{H}}_{rd,k}^{\mathrm{H}} \hat{\mathbf{H}}_{rd,k} (P_{r,k}\mathbf{\Psi}_{rd,k} + \sigma_{n_2}^2 \mathbf{I}_{N_R})^{-\frac{1}{2}} \\
&\approx \frac{1}{P_{r,k}} \mathbf{\Psi}_{rd,k}^{-\frac{\mathrm{H}}{2}} \hat{\mathbf{H}}_{rd,k}^{\mathrm{H}} \underbrace{\hat{\mathbf{H}}_{rd,k} \mathbf{\Psi}_{rd,k}^{-\frac{1}{2}}}_{\triangleq \mathbf{\Gamma}_k} \\
&= \frac{1}{P_{r,k}} \mathbf{U}_{\mathbf{\Gamma}_k} \mathbf{\Lambda}_{\mathbf{\Gamma}_k} \mathbf{U}_{\mathbf{\Gamma}_k}^{\mathrm{H}},
\end{aligned} \tag{68}$$

where the second equality is based on SVD. Comparing the right hand side of (68) with that of (38), we directly have

$$\mathbf{U}_{\mathbf{\Theta}_k} = \mathbf{U}_{\mathbf{\Gamma}_k}, \qquad \mathbf{\Lambda}_{\mathbf{\Theta}_k} = \frac{1}{P_{r,k}}\mathbf{\Lambda}_{\mathbf{\Gamma}_k}. \tag{69}$$

Substituting (69) into (51), γ_k reduces to a simpler form

$$\gamma_k = \frac{1}{P_{r,k}^2} \underbrace{\frac{c_{3,k}\sigma_{n_2}^2(c_{1,k} + c_{1,k}c_{4,k} - c_{2,k}c_{3,k})}{(1+c_{4,k})^2}}_{\triangleq \chi_k}, \tag{70}$$

where $c_{1,k}$, $c_{2,k}$, $c_{3,k}$ and $c_{4,k}$ are defined as

$$c_{1,k} \triangleq \mathrm{Tr}(\mathbf{U}_{\mathbf{\Gamma}_k,N_k}^{\mathrm{H}} \mathbf{\Psi}_{rd,k}^{-1} \mathbf{U}_{\mathbf{\Gamma}_k,N_k} \tilde{\mathbf{\Lambda}}_{\mathbf{T}_k} \tilde{\mathbf{\Lambda}}_{\mathbf{\Gamma}_k}^{-\frac{1}{2}} \mathbf{\Lambda}_{\mathbf{I},k}), \tag{71a}$$

$$c_{2,k} \triangleq \mathrm{Tr}(\mathbf{U}_{\mathbf{\Gamma}_k,N_k}^{\mathrm{H}} \mathbf{\Psi}_{rd,k}^{-1} \mathbf{U}_{\mathbf{\Gamma}_k,N_k} \tilde{\mathbf{\Lambda}}_{\mathbf{\Gamma}_k}^{-1} \mathbf{\Lambda}_{\mathbf{I},k}), \tag{71b}$$

$$c_{3,k} \triangleq \mathrm{Tr}(\tilde{\mathbf{\Lambda}}_{\mathbf{T}_k} \tilde{\mathbf{\Lambda}}_{\mathbf{\Gamma}_k}^{-\frac{1}{2}} \mathbf{\Lambda}_{\mathbf{I},k}), \tag{71c}$$

$$c_{4,k} \triangleq \mathrm{Tr}(\tilde{\mathbf{\Lambda}}_{\mathbf{\Gamma}_k}^{-1} \mathbf{\Lambda}_{\mathbf{I},k}), \tag{71d}$$

with the diagonal matrix $\tilde{\mathbf{\Lambda}}_{\mathbf{\Gamma}_k}$ being the $N_k \times N_k$ principal submatrix of $\mathbf{\Lambda}_{\mathbf{\Gamma}_k}$, and the matrix $\mathbf{U}_{\mathbf{\Gamma}_k,N_k}$ consists of the first N_k columns of the matrix $\mathbf{U}_{\mathbf{\Gamma}_k}$. Together with the facts that $\sum_k P_{r,k} = P_r$ and $\gamma_0 = \cdots = \gamma_{K-1}$, $P_{r,k}$ can be solved as

$$P_{r,k} = P_r \frac{\sqrt{\chi_k}}{\sum_k \sqrt{\chi_k}}. \tag{72}$$

2) Spectral Approximation (SPA): With SPA, the left hand side of (38) becomes

$$\begin{aligned} &(P_{r,k}\mathbf{\Psi}_{rd,k} + \sigma_{n_2}^2 \mathbf{I}_{N_R})^{-\frac{\mathrm{H}}{2}} \hat{\mathbf{H}}_{rd,k}^{\mathrm{H}} \hat{\mathbf{H}}_{rd,k} (P_{r,k}\mathbf{\Psi}_{rd,k} + \sigma_{n_2}^2 \mathbf{I}_{N_R})^{-\frac{1}{2}} \\ &\approx \frac{1}{P_{r,k}\lambda_{\max}(\mathbf{\Psi}_{rd,k}) + \sigma_{n_2}^2} \hat{\mathbf{H}}_{rd,k}^{\mathrm{H}} \hat{\mathbf{H}}_{rd,k}. \end{aligned} \tag{73}$$

Comparing (73) to (61), it is obvious that the problem becomes exactly the same as that discussed for uncorrelated channel estimation errors. Therefore, the allocated power to the k^{th} subcarrier $P_{r,k}$ can be calculated by (64) but with $\delta_{rd,k}$ replaced by $\lambda_{\max}(\mathbf{\Psi}_{rd,k})$.

Notice that the selection criterion $P_{r,k}\lambda_{\min}(\mathbf{\Psi}_{rd,k}) \gg \sigma_{n_2}^2$ involves the parameter of interest $P_{r,k}$. In practice, we can replace $P_{r,k}$ in the criterion by P_r/K. Then the proposed algorithm can be summarized as using HSA when $(P_r/K)\lambda_{\min}(\mathbf{\Psi}_{rd,k})/\sigma_{n_2}^2 \geq \mathcal{T}$, otherwise we use SPA, where $\mathcal{T}$ is a threshold.

Remark 5: Our design is valid for both cases whether the source has channel state information (CSI) or not, since our design is suitable for any correlation matrix $\mathbf{R}_{s,k}$ which is determined by the precoder at the source. The precoder at the source can be easily designed based on different criteria such as zero forcing (ZF), capacity maximization (CM) and minimum-mean-square-error (MMSE) using the first hop channel information. Once the precoder at the source is fixed, the optimal forward matrix at the relay and the equalizer at destination can be designed using the proposed design.

VI. Simulation Results and Discussions

In this section, we investigate the performance of the proposed algorithms. For the purpose of comparison, the algorithm based on estimated channel only (without taking the channel errors into account) is also simulated. An AF MIMO-OFDM relay system where the source, relay and destination are equipped with same number of antennas, $N_S = M_R = N_R = M_D = 2$ is considered. The number of subcarriers K is set to be 64, and the length of the multi-path channels in both hops is 5, and is denoted as L. The channel response is generated according to the HIPERLAN/2 standard [10]. The signal-to-noise ratio (SNR) of the first hop is defined as $\mathrm{E}_s/\mathrm{N}_1 = \mathrm{Tr}(\mathbf{R}_s)/(KM_R\sigma_{n_1}^2)$, and is fixed as $30\mathrm{dB}$. At the source, on each subcarrier, two independent data streams are transmitted by two antennas at the same power, and QPSK is used as the modulation scheme. The SNR at the second hop is defined as $\mathrm{E}_r/\mathrm{N}_2 = P_r/(KM_D\sigma_{n_2}^2)$. In the figures, MSE is referred to total simulated MSE over all subcarriers normalized by K.

Based on the definition of $\mathbf{D}$ in (9), $\mathbf{D}\mathbf{D}^{\mathrm{H}}$ is a block circular matrix. In the following, only the effect of spatial correlation of training sequence is demonstrated, and the training is white in time dimension. In this case, $\mathbf{D}\mathbf{D}^{\mathrm{H}}$ is a block diagonal matrix, and can be written as $\mathbf{D}\mathbf{D}^{\mathrm{H}} = \mathbf{I}_L \otimes \sum_i \mathbf{d}_i\mathbf{d}_i^{\mathrm{H}}$, where $\sum_i \mathbf{d}_i\mathbf{d}_i^{\mathrm{H}}/K$ is the spatial correlation matrix of the training sequence. Furthermore, the widely used exponential correlation model is adopted to denote the spatial correlation matrix [22], [23], and therefore we have

$$\mathbf{D}\mathbf{D}^{\mathrm{H}} = \mathbf{I}_L \otimes K \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix}. \tag{74}$$

It is assumed that the same training sequence is used for channel estimation in the two hops.

Based on the definition of $\boldsymbol{\Psi}_{sr,k}$ and $\boldsymbol{\Psi}_{rd,k}$ in (24) and (26), and together with (74), we have

$$\boldsymbol{\Psi}_{sr,k} = \boldsymbol{\Psi}_{rd,k} = \frac{L}{K}\sigma_e^2 \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix}^{-1}, \tag{75}$$

where σ_e^2 is the noise variance during channel estimation.

First, we investigate the performance of the proposed robust algorithm when channel estimation errors are uncorrelated, which corresponds to $\alpha = 0$ in (75). Fig. 2 shows the MSE of the received signal at the destination with different σ_e^2. It can be seen that the performance of the proposed robust algorithm is always better than that of the algorithm based on estimated CSI only, as long as σ_e^2 is not zero. Furthermore, the performance improvement of the proposed robust algorithm over the algorithm based on only estimated CSI enlarges when σ_e^2 increases. Fig. 3 shows the corresponding performance under correlated channel estimation errors ($\alpha = 0.4$), and the threshold $\mathcal{T}$ is set to be 10. It can be seen that a similar conclusion can be drawn as in Fig. 2.

Fig. 4 shows the MSE of the received signal at the destination for HSA, SPA and the proposed algorithm that switches between the two, when $\alpha = 0.4$. It is clear that HSA performs better than SPA at high SNR region, as at high SNR, (66) in HSA becomes equality. On the other hand, SPA performs better than HSA at low SNR region. The proposed robust algorithm combines the benefits of both HSA and SPA.

Fig. 5 shows the MSE of the output data at the destination for both proposed robust algorithm and the algorithm based on estimated CSI only, with different α and $\mathrm{E}_r/\mathrm{N}_2 = 25\mathrm{dB}$. It can be seen that although performance degradation is observed for both algorithms when α increases, the proposed robust algorithm shows a significant improvement over the algorithm based on estimated CSI only. Furthermore, as $\alpha = 0$ gives the best data MSE performance, it demonstrates that white sequence is prefered in channel estimation.

Finally, Fig. 6 shows the bit error rates (BER) of the output data at the destination for different σ_e^2, when $\alpha = 0.5$. It can be seen that the BER performance is consistent with MSE performance in Fig. 2 and Fig. 3.

VII. Conclusions

In this paper, linear robust relay precoder and destination equalizer were jointly designed for AF MIMO-OFDM relay systems based on MMSE criterion. The linear channel estimators and the corresponding MSE expressions were first derived. Then a general solution for optimal precoder and equalizer was proposed. When the channel estimation errors are uncorrelated, the optimal solution is in closed-form, and it includes several existing transceiver design results as special cases. On the other hand, when channel estimation errors are correlated, a practical algorithm was introduced. Simulation results showed that the proposed algorithms offer significant performance improvements over the algorithm based on estimated CSI only.

APPENDIX A
PROOF OF (7)

Based on the characteristics of DFT operation, the matrix $\boldsymbol{\mathcal{H}}_{sr}$ defined in (6) is a $KM_R \times KN_S$ block circulant matrix

$$\boldsymbol{\mathcal{H}}_{sr} \triangleq \begin{bmatrix} \boldsymbol{\mathcal{H}}_{sr}^{(0)} & \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\mathcal{H}}_{sr}^{(L_1-1)} & \boldsymbol{\mathcal{H}}_{sr}^{(L_1-2)} & \cdots & \boldsymbol{\mathcal{H}}_{sr}^{(1)} \\ \boldsymbol{\mathcal{H}}_{sr}^{(1)} & \boldsymbol{\mathcal{H}}_{sr}^{(0)} & \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\mathcal{H}}_{sr}^{(L_1-1)} & \cdots & \boldsymbol{\mathcal{H}}_{sr}^{(2)} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\mathcal{H}}_{sr}^{(L_1-1)} & \boldsymbol{\mathcal{H}}_{sr}^{(L_1-2)} & \boldsymbol{\mathcal{H}}_{sr}^{(L_1-3)} & \cdots & \boldsymbol{\mathcal{H}}_{sr}^{(0)} \end{bmatrix}, \tag{76}$$

where the element $\boldsymbol{\mathcal{H}}_{sr}^{(\ell)}$ is defined in (8). It is obvious that $\boldsymbol{\mathcal{H}}_{sr}^{(\ell)}$ is the ℓ^{th} tap of the multi-path MIMO channels between the source and relay in the time domain and L_1 is the length of the multi-path channel.

On the other hand, based on the definition of $\mathbf{d}$ in (6), we have

$$\mathbf{d} = [\underbrace{(\frac{1}{\sqrt{K}}\sum_{k=0}^{K-1}\mathbf{s}_k e^{j\frac{2\pi}{K}k(0)})^{\mathrm{T}}}_{\mathbf{d_0}} \ \underbrace{(\frac{1}{\sqrt{K}}\sum_{k=0}^{K-1}\mathbf{s}_k e^{j\frac{2\pi}{K}k(1)})^{\mathrm{T}}}_{\mathbf{d_1}} \ \cdots \ \underbrace{(\frac{1}{\sqrt{K}}\sum_{k=0}^{K-1}\mathbf{s}_k e^{j\frac{2\pi}{K}k(K-1)})^{\mathrm{T}}}_{\mathbf{d}_{K-1}}]^{\mathrm{T}}. \tag{77}$$

From (76) and (77), by straightforward computation, the signal model given in (6) can be reformulated as

$$\begin{aligned} \mathbf{r} = \boldsymbol{\mathcal{H}}_{sr}\mathbf{d} + \mathbf{v} &= \mathrm{vec}([\boldsymbol{\mathcal{H}}_{sr}^{(0)} \ \cdots \ \boldsymbol{\mathcal{H}}_{sr}^{(L_1-1)}]\mathbf{D}) + \mathbf{v} \\ &= (\mathbf{D}^{\mathrm{T}} \otimes \mathbf{I}_{M_R})\mathrm{vec}([\boldsymbol{\mathcal{H}}_{sr}^{(0)} \ \cdots \ \boldsymbol{\mathcal{H}}_{sr}^{(L_1-1)}]) + \mathbf{v}, \end{aligned} \tag{78}$$

where the matrix $\mathbf{D}$ is defined in (9).

APPENDIX B
PROOF OF (17)

For the expectation of the following product

$$\boldsymbol{\Sigma} = \mathbb{E}\{\mathbf{Q}\mathbf{R}\mathbf{W}^{\mathrm{H}}\} \tag{79}$$

where $\mathbf{Q}$ and $\mathbf{W}$ are two $M \times N$ random matrices with compatible dimension to $\mathbf{R}$, the $(i,j)^{\text{th}}$ element of $\boldsymbol{\Sigma}$ is

$$\boldsymbol{\Sigma}(i,j) = \mathbb{E}\{\mathbf{Q}(i,:)\mathbf{R}\mathbf{W}(j,:)^{\mathrm{H}}\} = \sum_t \sum_k \mathbb{E}\{\mathbf{Q}(i,t)\mathbf{R}(t,k)\mathbf{W}(j,k)^*\}. \tag{80}$$

If the two random matrices $\mathbf{Q}$ and $\mathbf{W}$ satisfy

$$\mathbb{E}\{\mathrm{vec}(\mathbf{Q})\mathrm{vec}^{\mathrm{H}}(\mathbf{W})\} = \mathbf{A} \otimes \mathbf{B}, \tag{81}$$

where $\mathbf{A}$ is a $N\times N$ matrix while $\mathbf{B}$ is a $M\times M$ matrix, then we have the equality $\mathbb{E}\{\mathbf{Q}(i_1,j_1)\mathbf{W}(i_2,j_2)^*\} = \mathbf{B}(i_1,i_2)\mathbf{A}(j_1,j_2)$. As $\mathbf{Q}(i,t)$ and $\mathbf{W}(j,k)$ are scalars, (80) can be further written as

$$\boldsymbol{\Sigma}(i,j) = \sum_t \sum_k (\mathbf{R}(t,k)\mathbb{E}\{\mathbf{Q}(i,t)\mathbf{W}(j,k)^*\}) = \sum_t \sum_k \mathbf{R}(t,k)\mathbf{A}(t,k)\mathbf{B}(i,j). \tag{82}$$

Finally, writing (82) back to matrix form, we have [36]

$$\boldsymbol{\Sigma} = \mathbf{B}\mathrm{Tr}(\mathbf{R}\mathbf{A}^{\mathrm{T}}). \tag{83}$$

Notice that this conclusion is independent of the matrix variate distributions of $\mathbf{Q}$ and $\mathbf{W}$, but only determined by their second order moments. Putting $\mathbf{A} = \sum_{\ell_2=0}^{L_1-1}\sum_{\ell_1=0}^{L_1-1}(e^{-j\frac{2\pi}{K}k(\ell_1-\ell_2)}\boldsymbol{\Phi}_{\ell_1,\ell_2}^{sr})$, $\mathbf{B} = \mathbf{I}_{M_R}$ and $\mathbf{Q} = \mathbf{W} = \Delta\mathbf{H}_{sr,k}$, into (83), we have (17).

APPENDIX C
PROOF OF PROPERTY 1

Right multiplying both sides of (34a) with $\mathbf{G}_k^{\mathrm{H}}$, the following equality holds

$$\mathbf{G}_k(\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\mathbf{F}_k^{\mathrm{H}}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}} + \mathbf{K}_k)\mathbf{G}_k^{\mathrm{H}} = \mathbf{R}_{s_k}(\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\hat{\mathbf{H}}_{sr,k})^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}}. \tag{84}$$

Left multiplying (34b) with $\mathbf{F}_k^{\mathrm{H}}$, we have

$$\begin{aligned}&\mathbf{F}_k^{\mathrm{H}}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}}\mathbf{G}_k\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k} + \mathbf{F}_k^{\mathrm{H}}\mathrm{Tr}(\mathbf{G}_k\mathbf{G}_k^{\mathrm{H}})\boldsymbol{\Psi}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k} + \gamma_k\mathbf{F}_k^{\mathrm{H}}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\\ &= \mathbf{F}_k^{\mathrm{H}}\left(\hat{\mathbf{H}}_{sr,k}\mathbf{R}_{s_k}\mathbf{G}_k\hat{\mathbf{H}}_{rd,k}\right)^{\mathrm{H}}.\end{aligned} \tag{85}$$

After taking the traces of both sides of (84) and (85) and with the fact that the traces of their righthand sides are equivalent, i.e., $\mathrm{Tr}(\mathbf{R}_{s_k}(\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\hat{\mathbf{H}}_{sr,k})^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}}) = \mathrm{Tr}(\mathbf{F}_k^{\mathrm{H}}(\hat{\mathbf{H}}_{sr,k}\mathbf{R}_{s_k}\mathbf{G}_k\hat{\mathbf{H}}_{rd,k})^{\mathrm{H}})$, we directly have

$$\begin{aligned}&\mathrm{Tr}(\mathbf{G}_k(\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\mathbf{F}_k^{\mathrm{H}}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}} + \mathbf{K}_k)\mathbf{G}_k^{\mathrm{H}}) \\ &= \mathrm{Tr}(\mathbf{F}_k^{\mathrm{H}}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}}\mathbf{G}_k\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}) + \gamma_k\mathrm{Tr}(\mathbf{F}_k^{\mathrm{H}}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}) + \mathrm{Tr}(\mathbf{G}_k\mathbf{G}_k^{\mathrm{H}})\mathrm{Tr}(\mathbf{F}_k^{\mathrm{H}}\mathbf{\Psi}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}).\end{aligned} \tag{86}$$

By the property of trace operator, $\mathrm{Tr}(\mathbf{G}_k(\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\mathbf{F}_k^{\mathrm{H}}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}})\mathbf{G}_k^{\mathrm{H}}) = \mathrm{Tr}(\mathbf{F}_k^{\mathrm{H}}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}}\mathbf{G}_k\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k})$, and (86) reduces to

$$\mathrm{Tr}(\mathbf{G}_k\mathbf{K}_k\mathbf{G}_k^{\mathrm{H}}) = \mathrm{Tr}(\mathbf{G}_k\mathbf{G}_k^{\mathrm{H}})\mathrm{Tr}(\mathbf{F}_k^{\mathrm{H}}\mathbf{\Psi}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}) + \gamma_k\mathrm{Tr}(\mathbf{F}_k^{\mathrm{H}}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}). \tag{87}$$

On the other hand, based on the definition of $\mathbf{K}_k$ in (30), $\mathrm{Tr}(\mathbf{G}_k\mathbf{K}_k\mathbf{G}_k^{\mathrm{H}})$ can be also expressed as

$$\mathrm{Tr}(\mathbf{G}_k\mathbf{K}_k\mathbf{G}_k^{\mathrm{H}}) = \mathrm{Tr}(\mathbf{G}_k\mathbf{G}_k^{\mathrm{H}})\mathrm{Tr}(\mathbf{F}_k^{\mathrm{H}}\mathbf{\Psi}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}) + \mathrm{Tr}(\mathbf{G}_k\mathbf{R}_{n_2,k}\mathbf{G}_k^{\mathrm{H}}). \tag{88}$$

Comparing (87) with (88), it can be concluded that

$$\mathrm{Tr}(\mathbf{G}_k\mathbf{R}_{n_2,k}\mathbf{G}_k^{\mathrm{H}}) = \gamma_k\mathrm{Tr}(\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\mathbf{F}_k^{\mathrm{H}}). \tag{89}$$

Putting (89) into (34c), we have $\mathrm{Tr}(\mathbf{G}_k\mathbf{R}_{n_2,k}\mathbf{G}_k^{\mathrm{H}}) - \gamma_k P_{r,k} = 0$. As $\mathbf{R}_{n_2,k} = \sigma_{n_2}^2\mathbf{I}_{M_D}$, it is straightforward that

$$\sigma_{n_2}^2\mathrm{Tr}(\mathbf{G}_k\mathbf{G}_k^{\mathrm{H}}) = \gamma_k P_{r,k}. \tag{90}$$

Furthermore, based on the fact $\gamma_0 = \gamma_1 = \cdots = \gamma_{K-1} = \gamma$ and taking summation of both sides of (90), the following equation holds

$$\sum_k \sigma_{n_2}^2\mathrm{Tr}(\mathbf{G}_k\mathbf{G}_k^{\mathrm{H}}) = \gamma\sum_k P_{r,k}. \tag{91}$$

Putting (91) into (34e), we have

$$\sum_k \sigma_{n_2}^2\mathrm{Tr}(\mathbf{G}_k\mathbf{G}_k^{\mathrm{H}}) - \gamma P_r = 0, \tag{92}$$

and it follows that

$$\gamma_k = \gamma = \sigma_{n_2}^2\frac{\sum_k \mathrm{Tr}(\mathbf{G}_k\mathbf{G}_k^{\mathrm{H}})}{P_r}. \tag{93}$$

Since for the optimal equalizer $\mathbf{G}_k$, $\sum_k \mathrm{Tr}(\mathbf{G}_{k,\mathrm{opt}}\mathbf{G}_{k,\mathrm{opt}}^{\mathrm{H}}) \neq 0$, it can be concluded that $\gamma_k \neq 0$. In order to have (34c) satisfied, we must have

$$\mathrm{Tr}(\mathbf{F}_{k,\mathrm{opt}}\mathbf{R}_{\mathbf{x}_k}\mathbf{F}_{k,\mathrm{opt}}^{\mathrm{H}}) = P_{r,k}. \tag{94}$$

Furthermore, as $\gamma \neq 0$, based on (34e), it is also concluded that

$$\sum_k P_{r,k} = P_r. \tag{95}$$

Finally, (90) constitutes the second part of the Property 1.

Appendix D
Proof of Property 2

Defining a full rank Hermitian matrix $\mathbf{M}_k = P_{r,k}\mathbf{\Psi}_{rd,k} + \sigma_{n_2}^2\mathbf{I}_{N_R}$, then for an arbitrary $N_R \times N_R$ matrix $\mathbf{F}_k$, it can be written as

$$\mathbf{F}_k = \mathbf{M}_k^{-\frac{1}{2}}\mathbf{U}_{\mathbf{\Theta}_k}\mathbf{\Sigma}_{\mathbf{F}_k}\mathbf{U}_{\mathbf{T}_k}^{\mathrm{H}}\mathbf{R}_{\mathbf{x}_k}^{-\frac{1}{2}} \tag{96}$$

where the inner matrix $\mathbf{\Sigma}_{\mathbf{F}_k}$ equals to $\mathbf{\Sigma}_{\mathbf{F}_k} = \mathbf{U}_{\mathbf{\Theta}_k}^{\mathrm{H}}\mathbf{M}_k^{\frac{1}{2}}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}^{\frac{1}{2}}\mathbf{U}_{\mathbf{T}_k}$.

Putting (96) into (34a), and with the following definitions (the same as the definitions in (38) and (39))

$$\mathbf{M}_k^{-\frac{\mathrm{H}}{2}}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}}\hat{\mathbf{H}}_{rd,k}\mathbf{M}_k^{-\frac{1}{2}} = \mathbf{U}_{\mathbf{\Theta}_k}\mathbf{\Lambda}_{\mathbf{\Theta}_k}\mathbf{U}_{\mathbf{\Theta}_k}^{\mathrm{H}}, \tag{97}$$

$$\mathbf{R}_{\mathbf{x},k}^{-\frac{1}{2}}\hat{\mathbf{H}}_{sr,k}\mathbf{R}_{s,k} = \mathbf{U}_{\mathbf{T}_k}\mathbf{\Lambda}_{\mathbf{T}_k}\mathbf{V}_{\mathbf{T}_k}^{\mathrm{H}}, \tag{98}$$

the equalizer $\mathbf{G}_k$ can be reformulated as

$$\begin{aligned}
\mathbf{G}_k &= \mathbf{R}_{\mathbf{s}_k}(\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\hat{\mathbf{H}}_{sr,k})^{\mathrm{H}}(\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\mathbf{F}_k^{\mathrm{H}}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}} + \eta_k\mathbf{I}_{M_D})^{-1} \\
&= (\mathbf{R}_{\mathbf{x}_k}^{-\frac{1}{2}}\hat{\mathbf{H}}_{sr,k}\mathbf{R}_{\mathbf{s}_k})^{\mathrm{H}}(\mathbf{R}_{\mathbf{x}_k}^{\frac{1}{2}}\mathbf{F}_k^{\mathrm{H}}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}}\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}^{\frac{1}{2}} + \eta_k\mathbf{I}_{M_R})^{-1}\mathbf{R}_{\mathbf{x}_k}^{\frac{1}{2}}\mathbf{F}_k^{\mathrm{H}}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}} \\
&= \mathbf{V}_{\mathbf{T}_k}\underbrace{\mathbf{\Lambda}_{\mathbf{T}_k}^{\mathrm{H}}(\mathbf{\Sigma}_{\mathbf{F}_k}^{\mathrm{H}}\mathbf{\Lambda}_{\mathbf{\Theta}_k}\mathbf{\Sigma}_{\mathbf{F}_k} + \eta_k\mathbf{I}_{M_R})^{-1}\mathbf{\Sigma}_{\mathbf{F}_k}^{\mathrm{H}}}_{\triangleq\mathbf{\Sigma}_{\mathbf{G}_k}}\mathbf{U}_{\mathbf{\Theta}_k}^{\mathrm{H}}\mathbf{M}_k^{-\frac{\mathrm{H}}{2}}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}},
\end{aligned} \tag{99}$$

where the second equality is due to the matrix inversion lemma.

Putting (90) from Appendix C into (34b), after multiplying both sides of (34b) with $\mathbf{M}_k^{-\frac{1}{2}}$, we have

$$\mathbf{M}_k^{-\frac{1}{2}}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}}\mathbf{G}_k\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}^{\frac{1}{2}} + \mathbf{M}_k^{\frac{1}{2}}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}^{\frac{1}{2}}\frac{\gamma_k}{\sigma_{n_2}^2} = \mathbf{M}_k^{-\frac{1}{2}}\left(\hat{\mathbf{H}}_{sr,k}\mathbf{R}_{s,k}\mathbf{G}_k\hat{\mathbf{H}}_{rd,k}\right)^{\mathrm{H}}\mathbf{R}_{\mathbf{x}_k}^{-\frac{1}{2}}. \tag{100}$$

Then substituting $\mathbf{F}_k$ in (96) and $\mathbf{G}_k$ in (99) into (100), we have

$$\mathbf{\Sigma_F} = (\mathbf{\Lambda}_{\mathbf{\Theta}_k}\mathbf{\Sigma}_{\mathbf{G}_k}^{\mathrm{H}}\mathbf{\Sigma}_{\mathbf{G}_k}\mathbf{\Lambda}_{\mathbf{\Theta}_k} + \frac{\gamma_k}{\sigma_{n_2}^2}\mathbf{I}_{N_R})^{-1}(\mathbf{\Lambda}_{\mathbf{T}_k}\mathbf{\Sigma}_{\mathbf{G}_k}\mathbf{\Lambda}_{\mathbf{\Theta}_k})^{\mathrm{H}}. \tag{101}$$

Since $\mathbf{\Lambda}_{\mathbf{T}_k}$ and $\mathbf{\Lambda}_{\mathbf{\Theta}_k}$ are rectangular diagonal matrices (denoting their ranks by p_k and q_k respectively), based on (101), it can be concluded that $\mathbf{\Sigma}_{\mathbf{F}_k}$ has the following form

$$\mathbf{\Sigma}_{\mathbf{F}_k} = \begin{bmatrix} \boldsymbol{A}_{\mathbf{F}_k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{N_R \times M_R}, \tag{102}$$

where $\boldsymbol{A}_{\mathbf{F}_k}$ is of dimension $q_k \times p_k$ and to be determined. Furthermore, putting (102) into the definition of $\mathbf{\Sigma}_{\mathbf{G}_k}$ in (99), we have

$$\mathbf{\Sigma}_{\mathbf{G}_k} = \begin{bmatrix} \boldsymbol{A}_{\mathbf{G}_k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{N_S \times M_D}, \tag{103}$$

where $\boldsymbol{A}_{\mathbf{G}_k}$ is of dimension $p_k \times q_k$, and to be determined. Substituting (102) and (103) into (96) and (99), it can be concluded that

$$\mathbf{F}_k = (P_{r,k}\mathbf{\Psi}_{rd,k} + \sigma_{n_2}^2\mathbf{I}_{N_R})^{-\frac{1}{2}}\mathbf{U}_{\mathbf{\Theta}_k,q_k}\boldsymbol{A}_{\mathbf{F}_k}\mathbf{U}_{\mathbf{T}_k,p_k}^{\mathrm{H}}\mathbf{R}_{\mathbf{x}_k}^{-\frac{1}{2}}, \tag{104}$$

$$\mathbf{G}_k = \mathbf{V}_{\mathbf{T}_k,p_k}\boldsymbol{A}_{\mathbf{G}_k}\mathbf{U}_{\mathbf{\Theta}_k,q_k}^{\mathrm{H}}(P_{r,k}\mathbf{\Psi}_{rd,k} + \sigma_{n_2}^2\mathbf{I}_{N_R})^{-\frac{\mathrm{H}}{2}}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}}, \tag{105}$$

where

$$\boldsymbol{A}_{\mathbf{G}_k} = \bar{\mathbf{\Lambda}}_{\mathbf{T}_k}^{\mathrm{H}}(\boldsymbol{A}_{\mathbf{F}_k}^{\mathrm{H}}\bar{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}\boldsymbol{A}_{\mathbf{F}_k} + \eta_k\mathbf{I}_{p_k})^{-1}\boldsymbol{A}_{\mathbf{F}_k}^{\mathrm{H}}, \tag{106}$$

and $\bar{\mathbf{\Lambda}}_{\mathbf{T}_k}$ is the $p_k \times p_k$ principal submatrix of $\mathbf{\Lambda}_{\mathbf{T}_k}$.

Appendix E
Proof of Property 3

Taking the trace of both sides of (42) and (43), and noticing that the resultant two equations are the same, it is obvious that

$$\mathrm{Tr}(\boldsymbol{A}_{\mathbf{G}_k}\bar{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}\boldsymbol{A}_{\mathbf{G}_k}^{\mathrm{H}}) = \frac{\gamma_k}{\eta_k\sigma_{n_2}^2}\mathrm{Tr}(\boldsymbol{A}_{\mathbf{F}_k}^{\mathrm{H}}\boldsymbol{A}_{\mathbf{F}_k}). \tag{107}$$

On the other hand, substituting (105) into (90) in Appendix C, we have

$$\mathrm{Tr}(\boldsymbol{A}_{\mathbf{G}_k}\bar{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}\boldsymbol{A}_{\mathbf{G}_k}^{\mathrm{H}}) = \frac{\gamma_k}{\sigma_{n_2}^2}P_{r,k}. \tag{108}$$

Comparing (107) and (108), it follows that

$$\frac{1}{\eta_k}\mathrm{Tr}(\boldsymbol{A}_{\mathbf{F}_k}^{\mathrm{H}}\boldsymbol{A}_{\mathbf{F}_k}) = P_{r,k}. \tag{109}$$

For the objective function in the optimization problem (32), substituting (40) and (41) into the MSE expression in (28), the MSE on the k^{th} subcarrier can be written as

$$\begin{aligned}&\mathrm{MSE}_k(\mathbf{F}_k, \mathbf{G}_k)\\&= \mathrm{Tr}(\bar{\boldsymbol{\Lambda}}_{\mathbf{T}_k}^2(\frac{1}{\eta_k}\boldsymbol{A}_{\mathbf{F}_k}^{\mathrm{H}}\bar{\boldsymbol{\Lambda}}_{\boldsymbol{\Theta}_k}\boldsymbol{A}_{\mathbf{F}_k} + \mathbf{I}_{p_k})^{-1}) + \underbrace{\mathrm{Tr}(\mathbf{R}_{\mathbf{s}_k}) - \mathrm{Tr}(\mathbf{R}_{\mathbf{s}_k}\hat{\mathbf{H}}_{sr,k}^{\mathrm{H}}\mathbf{R}_{\mathbf{x}_k}^{-1}\hat{\mathbf{H}}_{sr,k}\mathbf{R}_{\mathbf{s}_k})}_{\triangleq c_k},\end{aligned} \tag{110}$$

where c_k is a constant part independent of $\mathbf{F}_k$. Therefore, based on (109) and (110), the optimization problem (32) becomes as

$$\begin{aligned}\min_{\boldsymbol{A}_{\mathbf{F}_k}} \quad & \sum_k \mathrm{Tr}(\bar{\boldsymbol{\Lambda}}_{\mathbf{T}_k}^2(\frac{1}{\eta_k}\boldsymbol{A}_{\mathbf{F}_k}^{\mathrm{H}}\bar{\boldsymbol{\Lambda}}_{\boldsymbol{\Theta}_k}\boldsymbol{A}_{\mathbf{F}_k} + \mathbf{I}_{p_k})^{-1}) + c_k\\ \mathrm{s.t.} \quad & \frac{1}{\eta_k}\mathrm{Tr}(\boldsymbol{A}_{\mathbf{F}_k}^{\mathrm{H}}\boldsymbol{A}_{\mathbf{F}_k}) = P_{r,k},\\ & \sum_k P_{r,k} = P_r.\end{aligned} \tag{111}$$

For any given $P_{r,k}$, then the optimization problem (111) can be decoupled into a collection of the following sub-optimization problems

$$\begin{aligned}\min_{\boldsymbol{A}_{\mathbf{F}_k}} \quad & \mathrm{Tr}(\bar{\boldsymbol{\Lambda}}_{\mathbf{T}_k}^2(\frac{1}{\eta_k}\boldsymbol{A}_{\mathbf{F}_k}^{\mathrm{H}}\bar{\boldsymbol{\Lambda}}_{\boldsymbol{\Theta}_k}\boldsymbol{A}_{\mathbf{F}_k} + \mathbf{I}_{p_k})^{-1})\\ \mathrm{s.t.} \quad & \frac{1}{\eta_k}\mathrm{Tr}(\boldsymbol{A}_{\mathbf{F}_k}^{\mathrm{H}}\boldsymbol{A}_{\mathbf{F}_k}) = P_{r,k},\end{aligned} \tag{112}$$

where the constant part c_k is neglected. For any two $M \times M$ positive semi-definite Hermitian matrices $\mathbf{A}$ and $\mathbf{B}$, we have $\mathrm{Tr}(\mathbf{AB}) \geq \sum_i \lambda_i(\mathbf{A})\lambda_{M-i+1}(\mathbf{B})$, where $\lambda_i(\mathbf{Z})$ denotes the i^{th} largest eigenvalue of the matrix $\mathbf{Z}$ [37]. Together with the fact that elements of the diagonal matrix $\tilde{\boldsymbol{\Lambda}}_{\mathbf{T}_k}$ are in decreasing order, the objective function of (112) is minimized, when $(\boldsymbol{A}_{\mathbf{F}_k}^{\mathrm{H}}\bar{\boldsymbol{\Lambda}}_{\boldsymbol{\Theta}_k}\boldsymbol{A}_{\mathbf{F}_k}/\eta_k + \mathbf{I}_{N_k})$ is a diagonal matrix with the diagonal elements in decreasing order. The objective function can be rewritten as

$$\begin{aligned}&\mathrm{Tr}(\bar{\boldsymbol{\Lambda}}_{\mathbf{T}_k}^2(\frac{1}{\eta_k}\boldsymbol{A}_{\mathbf{F}_k}^{\mathrm{H}}\bar{\boldsymbol{\Lambda}}_{\boldsymbol{\Theta}_k}\boldsymbol{A}_{\mathbf{F}_k} + \mathbf{I}_{N_k})^{-1})\\&= \mathrm{d}^{\mathrm{T}}(\bar{\boldsymbol{\Lambda}}_{\mathbf{T}_k}^2)\underbrace{\mathrm{d}((\frac{1}{\eta_k}\boldsymbol{A}_{\mathbf{F}_k}^{\mathrm{H}}\bar{\boldsymbol{\Lambda}}_{\boldsymbol{\Theta}_k}\boldsymbol{A}_{\mathbf{F}_k} + \mathbf{I}_{N_k})^{-1})}_{\triangleq \mathbf{b}} \triangleq \boldsymbol{f}(\mathbf{b}),\end{aligned} \tag{113}$$

where $\mathrm{d}(\mathbf{Z})$ denotes the vector which consists of the main diagonal elements of the matrix $\mathbf{Z}$.

It follows that $\boldsymbol{f}(\mathbf{b})$ is a schur-concave function of $\mathbf{b}$ [37, *3.H.3*]. Then, based on [15, *Theorem 1*], the optimal $\boldsymbol{A}_{\mathbf{F}_k}$ has the following structure

$$\boldsymbol{A}_{\mathbf{F}_k,\mathrm{opt}} = \begin{bmatrix} \boldsymbol{\Lambda}_{\mathbf{F}_k,\mathrm{opt}} & \mathbf{0}_{N_k,p_k-N_k} \\ \mathbf{0}_{q_k-N_k,N_k} & \mathbf{0}_{q_k-N_k,p_k-N_k} \end{bmatrix}, \tag{114}$$

where $\boldsymbol{\Lambda}_{\mathbf{F}_k,\mathrm{opt}}$ is a $N_k \times N_k$ diagonal matrix to be determined, and $N_k = \min(p_k, q_k)$.

Putting (114) into the definition of $\boldsymbol{A}_{\mathbf{G}_k,\mathrm{opt}}$ in (106), the structure of the optimal $\boldsymbol{A}_{\mathbf{G}_k,\mathrm{opt}}$ is given by

$$\boldsymbol{A}_{\mathbf{G}_k,\mathrm{opt}} = \begin{bmatrix} \boldsymbol{\Lambda}_{\mathbf{G}_k,\mathrm{opt}} & \mathbf{0}_{N_k,q_k-N_k} \\ \mathbf{0}_{p_k-N_k,N_k} & \mathbf{0}_{p_k-N_k,q_k-N_k} \end{bmatrix}, \tag{115}$$

where $\boldsymbol{\Lambda}_{\mathbf{G}_k,\mathrm{opt}}$ is also a $N_k \times N_k$ diagonal matrix.

REFERENCES

[1] A. Scaglione, D. L. Goeckel, and J. N. Laneman, "Cooperative communications in mobile Ad Hoc networks," *IEEE Signal Processing Magazine*, pp. 18–29, Sept. 2006.

[2] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Information Theory*, vol. 50, no. 12 pp. 3062–3080, Dec. 2004.

[3] O. Munoz-Medina, J. Vidal, and A. Agustin, "Linear transceiver design in nonregenerative relays with channel state information," *IEEE Trans. Signal Processing*, vol. 55, no. 6, pp. 2198–2206, Jun. 2007.

[4] X. Tang and Y. Hua, "Optimal design of non-regenerative MIMO wireless relays," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1398–1407, Apr. 2007.

[5] C.-B. Chae, T. W. Tang, R. W. Health, and S.-Y. Cho, "MIMO relaying with linear processing for multiuser transmission in fixed relay networks," *IEEE Trans. Signal Processing*, vol. 56, no. 2, pp. 727–738, Feb. 2008.

[6] A. S. Behbahani, R. Merched, and A. J. Eltawil, "Optimizations of a MIMO relay Network, " *IEEE Trans. Signal Processing*, vol. 56, no. 10, part 2, pp. 5062–5073, Oct. 2008.

[7] B. Khoshnevis, W. Yu and R. Adve, "Grassmannian beamforming for MIMO amplify-and-forward relaying," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1397-1407, Oct. 2008.

[8] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press. 2005

[9] S. Ma and T.-S. Ng,"Time domain signal detection based on second-order statistics for MIMO-OFDM system", *IEEE Transactions on Signal Processing*, vol. 55(3), pp. 1150-1158, Mar. 2007

[10] J. Chen, Y.-C. Wu, S. Ma and T.-S. Ng, "Joint CFO and channel estimation for multiuser MIMO-OFDM systems with optimal training sequences," *IEEE Trans. Signal Processing*, vol. 56, no. 8, pp. 4008–4019, Aug. 2008.

[11] F. Verde, D. Darsena and A. Scaglione, "Cooperative Randomized MIMO-OFDM Downlink for Multicell Networks: Design and Analysis," *IEEE Trans. Signal Processing*, vol. 58, no. 1, pp. 384–402, Jan. 2010.

[12] I. Hammerstrom and A. Wittneben, "Power allocation schemes for amplify-and-forward MIMO-OFDM relay links, " *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 2798–2802, Aug. 2007.

[13] H. Bolcskei, D. Gesbert, C. B. Papadias, and A.-J. Van Der Veen, *Space-Time Wireless Systems*. Cambridge University Press. 2006

[14] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2198–2206, Dec. 2001.

[15] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint robust Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. on Signal Processing*, vol. 51, no. 9, pp. 2381–2399, Sep. 2003.

[16] M. Joham, W. Utschick, and J. A. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 2700–2712, Aug. 2005.

[17] S. Serbetli and A. Yener, "Transceiver optimization for mutiuser MIMO systems," *IEEE Trans. on Signal Processing*, vol. 52, no. 1, pp. 214–226, Jan. 2004.

[18] Z. Q. Luo, T. N. Davidson, G. B. Giannakis and K. M. Wong, "Transceiver optimization for block-based multiple access through ISI channels," *IEEE Trans. on Signal Processing*, vol. 52, no. 4, pp. 1037–1052, Apr. 2004.

[19] W. Guan and H. Luo, "Joint MMSE transceiver design in non-regenerative MIMO relay systems," *IEEE Communications Lett.*, vol. 12, issue 7, pp. 517–519, July 2008.

[20] Y. Rong, X. Tang, and T. Hua, "A unified framework for optimizing linear non-regenerative multicarrier MIMO relay communication systems," *IEEE Trans. Signal Processing*, vol. 57, no. 12, pp. 4837–4851, Dec. 2009.

[21] L. Musavian, M. R. Nakhi, M. Dohler, and A. H. Aghvami, "Effect of channel uncertianty on the mutual information of MIMO fading channels," *IEEE Trans. Veh. Technol.*, vol. 56, no. 5, pp. 2798–2806, Sep. 2007.

[22] M. Ding and S. D. Blostein, "MIMO minimum total MSE transceiver design with imperfect CSI at both ends," *IEEE Trans. on Signal Processing*, vol. 57, no. 3, pp. 1141–1150, March 2009.

[23] X. Zhang, D. P. Palomar, and B. Ottersten, "Statistically robust design of linear MIMO transceiver," *IEEE Trans. Signal Processing*, vol. 56, no. 8, pp. 3678–3689, Aug. 2008.

[24] F. Rey, M. Lamarca, and G. Vazquez, "Robust power allocation algorithms for MIMO OFDM systems with imperfect CSI," *IEEE Trans. Signal Processing*, vol. 53, no. 3, pp. 1070–1085, Mar. 2005.

[25] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*. Cambridge University Press, 2003.

[26] C. P. Robert, *The Bayesian Choice*. New York: Springer Press, 2001.

[27] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[28] D. P. Bertsekas, and A. Nedic and A. E. Ozdaglar, *Convex Analysis and Optimization*. Athena Scientific, 2003.

[29] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press. 1985.

[30] P. Stoica and O. Besson, "Training sequence deisgn for frequency offset and frequency-selective channel estimation," *IEEE Trans. Commun.*, vol. 51, no. 11, pp.1910–1917, Nov. 2003.

[31] M. Ghogho and A. Swami, "Training design for multipath channel and frequency-Offset Estimation in MIMO systems," *IEEE Trans. Signal Processing*, vol. 54, no. 10, pp. 3957–3965, Oct. 2006.

[32] H. Minn and N. Al-Dhahir, "Optimal training signals for MIMO OFDM channel estimation," *IEEE Trans. Wireless Commun.*, vol. 5, no. 5, pp. 1158–1168, May 2006.

[33] A. Beck, A. Ben-Tal and Y. C. Eldar, "Robust mean-squared error estimation of multiple signals in linear systems affected by model and noise uncertainties", *Math. Programming*, vol.107 155-187, Springer, 2006.

[34] J. G. Andrews, A. Ghosh, and R. Muhamed, *Fundamentals of WiMAX*. USA: Prentice Hall, 2007.

[35] F. Ohrtman and K. Roeder, *Wi-Fi Handbook: Building 802.11 b wireless networks*. McGraw-Hill, 2003.

[36] S. Kay, *Fundamental of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[37] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic, 1979.

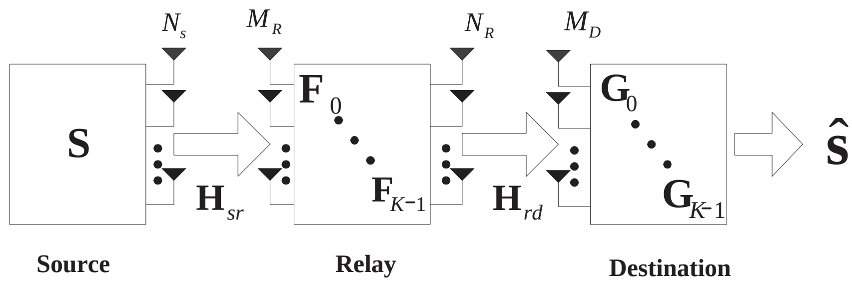

Fig. 1. Amplify-and-forward MIMO relaying.

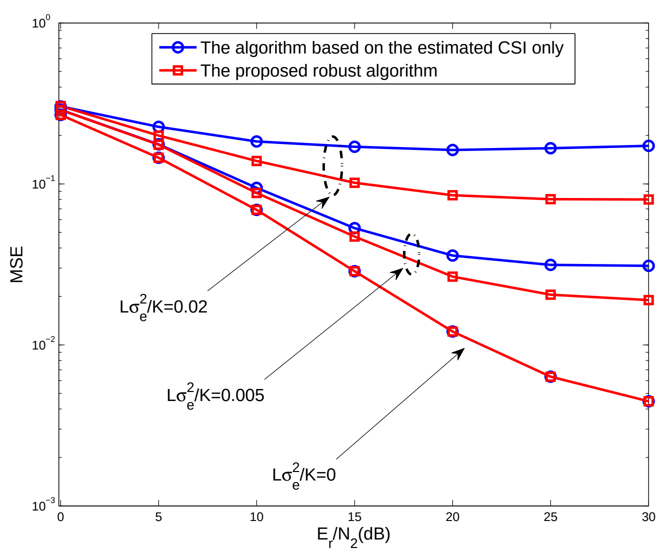

Fig. 2. MSE of received signal at the destination for different σ_e^2 when $\alpha = 0$.

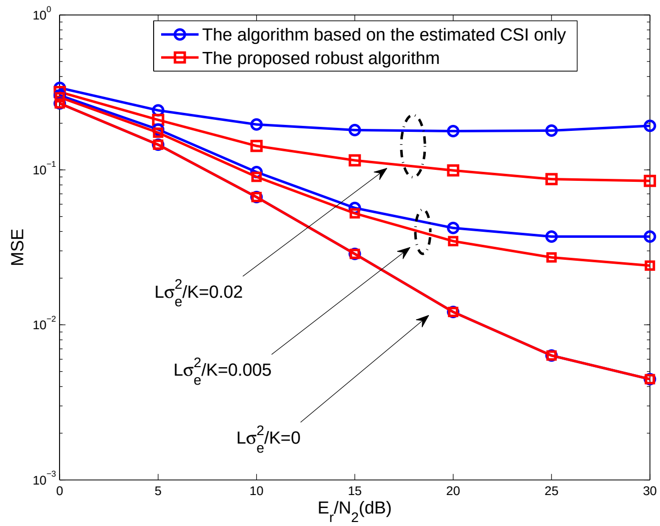

Fig. 3. MSE of received signal at the destination for different σ_e^2 when $\alpha = 0.4$.

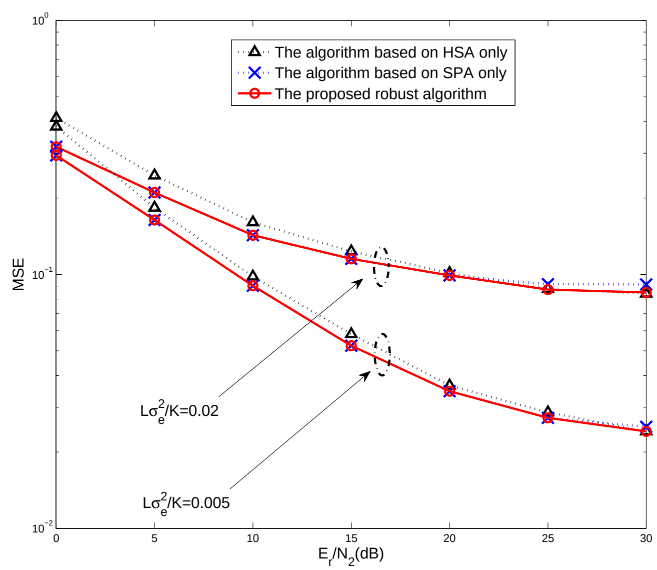

Fig. 4. MSE of received data at the destination for HSA, SPA and proposed robust algorithm when $\alpha = 0.4$.

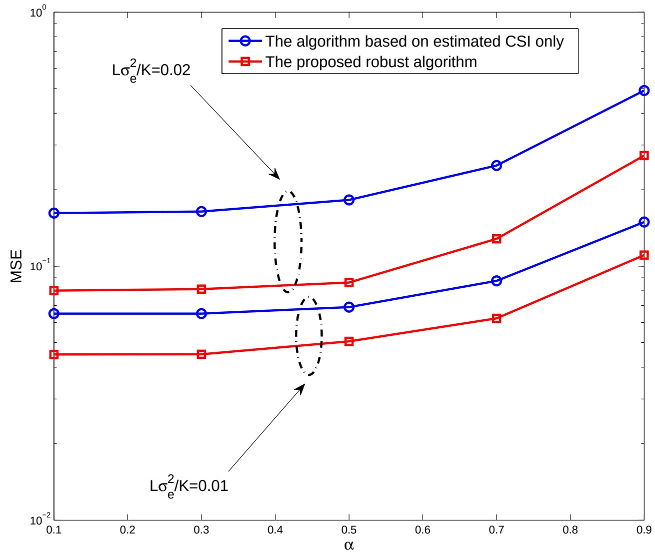

Fig. 5. MSE of received data at the destination for different α.

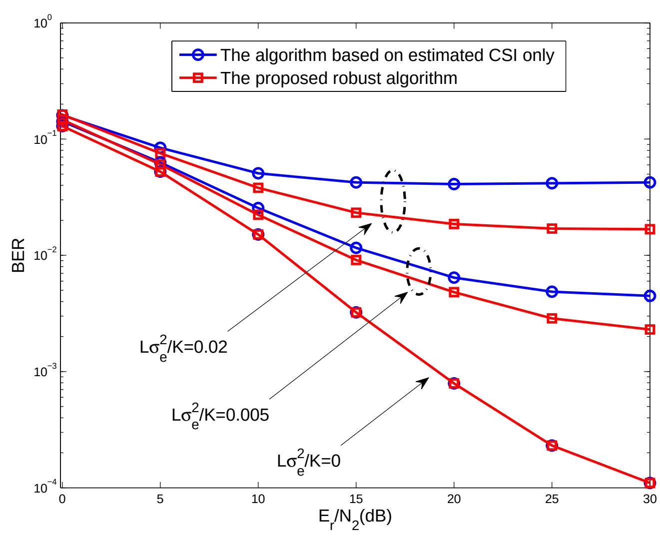

Fig. 6. BER of received data at the destination for different σ_e^2 when $\alpha = 0.5$