Transceiver Design for Dual-Hop Nonregenerative MIMO-OFDM Relay Systems Under Channel Uncertainties

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

***Abstract*—In this paper, linear transceiver design for dual-hop nonregenerative [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 forwarding matrix at the relay and equalizer at the destination under channel estimation errors 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. Moreover, this design is extended to the joint design involving source precoder design. Simulation results show that the proposed design outperforms the design based on estimated channel state information only.**

***Index Terms*—Amplify-and-forward (AF), equalizer, forwarding matrix, minimum mean-square-error (MMSE).**

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], [18]. For single carrier AF MIMO relay systems, linear MMSE forwarding matrix 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 based on prefect channel state information (CSI) is proposed in [20].

In all the above works, 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 designs can be employed: min-max and stochastic designs. If the distributions of channel estimation errors are known to be unbounded, stochastic design is preferred. Stochastic 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 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 linear MMSE transceiver design for dual-hop AF MIMO-OFDM relay systems without the 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 relay forwarding matrix and destination equalizer is proposed. Both the uncorrelated and correlated channel estimation errors are considered. The relationship between the proposed algorithm and several existing designs is revealed. Furthermore, the proposed closed-form solution is further extended to an iterative algorithm for joint design of source precoder, relay forwarding matrix and destination equalizer. Simulation results demonstrate that the

Manuscript received March 15, 2010; accepted August 15, 2010. Date of publication August 30, 2010; date of current version November 17, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Xiqi Gao.

C. Xing is with the School of Information and Electronics, Beijing Institute of Technology, Beijing, China (e-mail: cwxing@eee.hku.hk).

S. Ma, Y.-C. Wu, and T.-S. Ng are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong (e-mail: sdma@eee.hku.hk; ycwu@eee.hku.hk; tsng@eee.hku.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2010.2070797

proposed algorithms provide an obvious advantage in terms of data mean-square-error (MSE) compared to the algorithm based on estimated CSI only.

We want to highlight that the solution proposed in this paper can be directly extended to the problem minimizing the weighted MSE. Various objective metrics such as capacity maximization and minimizing maximum MSE can be transformed to a weighted MSE problem with different weighting matrices [14]. For clearness of presentation, we only consider a sum MSE minimization problem. On the other hand, minimizing the transmit power with a QoS requirement is a different perspective for transceiver design. Formulating and solving this problem is out of the scope of this paper.

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 relay forwarding matrix and destination equalizer design problem is proposed. The proposed closed-form solution is further extended to an iterative algorithm to include the design of source precoder in Section VI. Simulation results are given in Section VII and finally, conclusions are drawn in Section VIII.

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 $\mathbf{Z}^{1/2}$ is the Hermitian square root of the positive semidefinite matrix $\mathbf{Z}$, such that $\mathbf{Z} = \mathbf{Z}^{1/2}\mathbf{Z}^{1/2}$ and $\mathbf{Z}^{1/2}$ 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, we consider a dual-hop AF MIMO-OFDM relaying cooperative communication system, 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 $\mathbf{R}_{\mathbf{s}_k} = \mathbb{E}\{\mathbf{s}_k\mathbf{s}_k^{\mathrm{H}}\}$ 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 forwarding matrix $\mathbf{F}_k$, under a power constraint $\sum_k \mathrm{Tr}(\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\mathbf{F}_k^{\mathrm{H}}) \le P_r$ where $\mathbf{R}_{\mathbf{x}_k} = \mathbb{E}\{\mathbf{x}_k\mathbf{x}_k^{\mathrm{H}}\}$ 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

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

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

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

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

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

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

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 this paper, 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

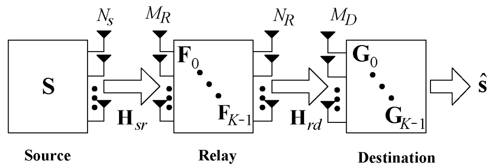


Fig. 1. AF MIMO-OFDM relaying diagram.

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

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

where $\boldsymbol{\mathcal{F}}$ is the normalized 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

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

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

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

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

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

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

$$\begin{aligned}\hat{\boldsymbol{\xi}}_{sr} = &\left(\sigma_{n_1}^{-2}\left(\mathbf{D}^{\mathrm{T}} \otimes \mathbf{I}_{M_R}\right)^{\mathrm{H}}\left(\mathbf{D}^{\mathrm{T}} \otimes \mathbf{I}_{M_R}\right) + \mathbf{R}_{\mathrm{channel}}^{-1}\right)^{-1}\\ &\times \sigma_{n_1}^{-2}\left(\mathbf{D}^{\mathrm{T}} \otimes \mathbf{I}_{M_R}\right)^{\mathrm{H}}\mathbf{y} \end{aligned} \tag{11}$$

with the corresponding MSE

$$\begin{aligned}&\mathbb{E}\left\{(\boldsymbol{\xi}_{sr} - \hat{\boldsymbol{\xi}}_{sr})(\boldsymbol{\xi}_{sr} - \hat{\boldsymbol{\xi}}_{sr})^{\mathrm{H}}\right\}\\ &\quad = \left(\mathbf{R}_{\mathrm{channel}}^{-1} + \sigma_{n_1}^{-2}(\mathbf{D}^*\mathbf{D}^{\mathrm{T}}) \otimes \mathbf{I}_{M_R}\right)^{-1} \end{aligned} \tag{12}$$

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

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

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

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

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

In case there is no prior information on $\mathbf{R}_{\mathrm{channel}}$, 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, $\mathbf{R}_{\mathrm{channel}}^{-1} \rightarrow \mathbf{0}$, 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

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

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

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

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

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:

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

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 forwarding 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:

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

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 $\mathbf{n}_{2,k}$.[2] Since $\mathbf{s}_k$, $\mathbf{n}_{1,k}$ and $\mathbf{n}_{2,k}$ are independent, the MSE expression (20) can be written as

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

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

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

[2]In this paper, the MSE is in fact an average of the traditional MSE over all possible channel estimation errors $\Delta\mathbf{H}_{sr,k}$ and $\Delta\mathbf{H}_{rd,k}$. When the LMMSE channel estimator is adopted, it is equivalent to the conditional MSE corresponding to the partial CSI case defined in [27].

For the inner expectation, the following equation holds:

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

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

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

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

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

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

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

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

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

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

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

where

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

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

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 relay forwarding matrix and destination equalizer that minimizes the total MSE of the output data at the destination can be formulated as the following optimization problem:

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

Remark 1: In this paper, the relay estimates the source-relay channel and the destination estimates the relay-destination channel. The forwarding matrix $\mathbf{F}_k$ and equalizer $\mathbf{G}_k$ are designed at the relay. Therefore, the estimated second hop CSI should be fed back from destination to relay. 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 for $\mathbf{G}_k$'s and $\mathbf{F}_k$'s

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

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

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

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

where the positive scalars γ_k and ρ 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 [28]

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

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

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

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

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

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

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

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

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 [29]. Based on these facts, the KKT conditions are the necessary conditions.[3] From KKT conditions, we can derive 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

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

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

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

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

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 eigenvalue decomposition (EVD) and singular value decomposition (SVD) as

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

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

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 forwarding matrix $\mathbf{F}_k$ and equalizer $\mathbf{G}_k$ must be in the form

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

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

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 $p_k = \mathrm{Rank}(\mathbf{\Lambda}_{\mathbf{T}_k})$. Similarly, $\mathbf{U}_{\mathbf{\Theta}_k,q_k}$ is the first q_k columns of $\mathbf{U}_{\mathbf{\Theta}_k}$, and $q_k = \mathrm{Rank}(\mathbf{\Lambda}_{\mathbf{\Theta}_k})$.

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

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

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

[3]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 the matrix $\bar{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}$ is the $q_k \times q_k$ principal submatrix of $\mathbf{\Lambda}_{\mathbf{\Theta}_k}$. Similarly, $\bar{\mathbf{\Lambda}}_{\mathbf{T}_k}$ is the $p_k \times p_k$ principal submatrix of $\mathbf{\Lambda}_{\mathbf{T}_k}$. 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 point-to-point systems [14], [22]. Then, the solutions satisfying KKT conditions and obtained by solving (42) and (43) are not unique. 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:

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

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

where $\mathbf{\Lambda}_{\mathbf{F}_k,\mathrm{opt}}$ and $\mathbf{\Lambda}_{\mathbf{G}_k,\mathrm{opt}}$ are two $N_k \times N_k$ diagonal matrices to be determined, and $N_k = \min(p_k, q_k)$. Notice that Property 3 is obtained by applying majorization theory to the original optimization problem. It is also a necessary condition for the optimal solution, and contains different information from that of Property 2.

Combining Property 2 and Property 3, and following the argument in [14], it can be concluded that the optimal solution of $\boldsymbol{A}_{\mathbf{F}_k}$ and $\boldsymbol{A}_{\mathbf{G}_k}$ is unique. Now, substituting (44) and (45) into (42) and (43), and noticing that all matrices are diagonal, $\mathbf{\Lambda}_{\mathbf{F}_k,\mathrm{opt}}$ and $\mathbf{\Lambda}_{\mathbf{G}_k,\mathrm{opt}}$ can be easily solved to be

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

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

where the matrices $\tilde{\mathbf{\Lambda}}_{\mathbf{T}_k}$ and $\tilde{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}$ are the principal submatrices of $\mathbf{\Lambda}_{\mathbf{T}_k}$ and $\mathbf{\Lambda}_{\mathbf{\Theta}_k}$ with dimension $N_k \times N_k$, and $N_k = \min\{\mathrm{rank}(\mathbf{\Lambda}_{\mathbf{\Theta}_k}), \mathrm{rank}(\mathbf{\Lambda}_{\mathbf{T}_k})\}$. 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. From (46) and (47), it can be seen that the optimal solutions are variants of water-filling solution. Furthermore, the eigen channels of two hops are paired based on the best-to-best criterion at the relay.

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 forwarding matrix and equalizer should simultaneously satisfy

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

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

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

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

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

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

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

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

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

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

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 forwarding matrix $\mathbf{F}_{k,\mathrm{opt}}$ and equalizer $\mathbf{G}_{k,\mathrm{opt}}$ are

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

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

where

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

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

with η_k and γ_k given by (50)–(52).

From the above summary, it is obvious that the problem of finding optimal forwarding matrix 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. Note that for single-carrier systems no power allocation has to be calculated since only one carrier exists, i.e., $P_{r,1} = P_r$, $K = 1$. In this case, the proposed closed-form solution is exactly the optimal solution for the transceiver design under channel estimation errors 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 linear MMSE transceiver under channel uncertainties 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 LMMSE transceiver under channel uncertainties [22].

Remark 4: The complexity of the proposed algorithm is dominated by one matrix inversion of $(P_{r,k}\mathbf{\Psi}_{rd,k}+\sigma_{n_2}^2\mathbf{I}_{N_R})^{-(1/2)}$, three matrix multiplications and one EVD in (38), one matrix inversion of $\mathbf{R}_{\mathbf{x}_k}^{-(1/2)}$, two matrix multiplications and one SVD in (39), four matrix multiplications in (53), four matrix multiplications in (54), and two water-filling computations in (55) and (56). Note that the matrix inversions in (53) and (54) are the same as those in (38) and (39) and therefore their computations could be saved. Specifically, in (38), the matrix inversion, matrix multiplications and EVD operation have complexities of $O({N_R}^3)$, $O(2{N_R}^3+{N_R}^2M_D)$ and $O({N_R}^3)$, respectively [30]. In (39), the matrix inversion, matrix multiplications and SVD operation costs $O({M_R}^3)$, $O({M_R}^2N_S+M_R{N_S}^2)$, and $O({M_R}^2N_S)$, respectively. With the diagonal structures of $\mathbf{\Lambda}_{\mathbf{F}_k,\mathrm{opt}}$ and $\mathbf{\Lambda}_{\mathbf{G_k},\mathrm{opt}}$, the matrix multiplications in (53) and (54) have complexities of $O({N_R}^2N_k+N_RN_k+N_RN_kM_R+{M_R}^2N_R)$ and $O(N_SN_k+N_SN_RN_k+{N_R}^2N_S+N_RN_SM_D)$, respectively. On the other hand, the complexities for the two water-filling computations in (55) and (56) are $O(N_k^2)$. As a result, for the AF MIMO-OFDM system with K subcarriers, the complexity of the proposed transceiver design is approximately upper bounded by $O(Km^3)$, where $m=\max\{M_D,N_R,M_R,N_S\}$.

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], [31]–[33]:

$$\mathbf{D}\mathbf{D}^{\mathrm{H}}\propto\mathbf{I}_{N_SL_1}. \tag{59}$$

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

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

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

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

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

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

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

$$\begin{aligned}P_{r,k}=&\sqrt{\frac{\sigma_{n_2}^2}{\rho}}\frac{\mathrm{Tr}\left(\tilde{\mathbf{\Lambda}}_{\mathbf{T}_k}\tilde{\mathbf{\Lambda}}_{\mathbf{H}_k}^{-\frac{1}{2}}\mathbf{\Lambda}_{\mathbf{I},k}\right)}{1+\delta_{rd,k}\mathrm{Tr}\left(\tilde{\mathbf{\Lambda}}_{\mathbf{H}_k}^{-1}\mathbf{\Lambda}_{\mathbf{I},k}\right)}\\&-\frac{\sigma_{n_2}^2\mathrm{Tr}\left(\tilde{\mathbf{\Lambda}}_{\mathbf{H}_k}^{-1}\mathbf{\Lambda}_{\mathbf{I},k}\right)}{1+\delta_{rd,k}\mathrm{Tr}\left(\tilde{\mathbf{\Lambda}}_{\mathbf{H}_k}^{-1}\mathbf{\Lambda}_{\mathbf{I},k}\right)},\quad k=0,\cdots K-1\end{aligned} \tag{64}$$

where ρ equals

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

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 [31]. 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. Here, we employ the spectral approximation (SPA)

$$P_{r,k}\mathbf{\Psi}_{rd,k}+\sigma_{n_2}^2\mathbf{I}_{N_R}\approx\left(P_{r,k}\lambda_{\max}(\mathbf{\Psi}_{rd,k})+\sigma_{n_2}^2\right)\mathbf{I}_{N_R}. \tag{66}$$

For spectral approximation, $\mathbf{\Psi}_{rd,k}$ is replaced by $\lambda_{\max}(\mathbf{\Psi}_{rd,k})\mathbf{I}$, where $\lambda_{\max}(\mathbf{\Psi}_{rd,k})$ is the maximum eigenvalue of $\mathbf{\Psi}_{rd,k}$. Applying (66) to 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 [35], [36], the eigenvalue spread of $\mathbf{\Psi}_{rd}$ is small, and SPA is a good approximation. With SPA, the left-hand side of (38) becomes

$$\begin{aligned}&\left(P_{r,k}\mathbf{\Psi}_{rd,k}+\sigma_{n_2}^2\mathbf{I}_{N_R}\right)^{-\frac{\mathrm{H}}{2}}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}}\hat{\mathbf{H}}_{rd,k}\\&\times\left(P_{r,k}\mathbf{\Psi}_{rd,k}+\sigma_{n_2}^2\mathbf{I}_{N_R}\right)^{-\frac{1}{2}}\\&\approx\frac{1}{P_{r,k}\lambda_{\max}(\mathbf{\Psi}_{rd,k})+\sigma_{n_2}^2}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}}\hat{\mathbf{H}}_{rd,k}.\end{aligned} \tag{67}$$

Comparing (67) 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 $\lambda_{\max}(\mathbf{\Psi}_{rd,k})$.

VI. Extension to the Joint Design Involving Source Precoder

Notice that the design in the previous section is suitable for scenarios where the source has fixed precoder. For example, the source precoder can be set to $\mathbf{I}$ for full spatial multiplexing or space-time block coding matrix for increasing diversity. On the other hand, if source precoder, relay forwarding matrix and destination equalizer are jointly designed, we can proceeds as follows. First, with a source precoder $\mathbf{P}_k$ before transmission, the system model in (2) is rewritten as

$$\mathbf{y}_k = \mathbf{H}_{rd,k}\mathbf{F}_k\mathbf{H}_{sr,k}\mathbf{P}_k\mathbf{s}_k + \mathbf{H}_{rd,k}\mathbf{F}_k\mathbf{n}_{1,k} + \mathbf{n}_{2,k}. \tag{68}$$

It can be seen that (68) is the same as (2) except $\mathbf{H}_{sr,k}\mathbf{P}_k$ is in the place of $\mathbf{H}_{sr,k}$. Furthermore, without loss of generality, we can assume $\mathbf{R}_{\mathbf{s}_k} = \mathbf{I}_{N_k}$ in (68) as all correlations are represented by $\mathbf{P}_k$. Then by using the substitutions $\mathbf{H}_{sr,k} \rightarrow \mathbf{H}_{sr,k}\mathbf{P}_k$ and $\mathbf{R}_{\mathbf{s}_k} \rightarrow \mathbf{I}_{N_k}$ into the first line of (21), and following the same derivation in Section IV, it can be easily proved that the data MSE at destination in the k^{th} subcarrier is

$$\begin{aligned}\text{MSE}_k&(\mathbf{G}_k, \mathbf{F}_k, \mathbf{P}_k)\\ &= \text{Tr}\left(\mathbf{G}_k\left(\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\mathbf{F}_k^{\text{H}}\hat{\mathbf{H}}_{rd,k}^{\text{H}} + \mathbf{K}_k\right)\mathbf{G}_k^{\text{H}}\right)\\ &\quad - \text{Tr}\left(\mathbf{P}_k^{\text{H}}\hat{\mathbf{H}}_{sr,k}^{\text{H}}\mathbf{F}_k^{\text{H}}\hat{\mathbf{H}}_{rd,k}^{\text{H}}\mathbf{G}_k^{\text{H}}\right)\\ &\quad - \text{Tr}\left(\mathbf{G}_k\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\hat{\mathbf{H}}_{sr,k}\mathbf{P}_k\right) + \text{Tr}\left(\mathbf{I}_{N_k}\right)\end{aligned} \tag{69}$$

where

$$\mathbf{R}_{\mathbf{x}_k} = \text{Tr}\left(\mathbf{P}_k\mathbf{P}_k^{\text{H}}\mathbf{\Psi}_{sr,k}\right)\mathbf{I}_{M_R} + \hat{\mathbf{H}}_{sr,k}\mathbf{P}_k\mathbf{P}_k^{\text{H}}\hat{\mathbf{H}}_{sr,k}^{\text{H}} + \sigma_{n_1}^2\mathbf{I}_{M_R}. \tag{70}$$

Comparing (28) to (69), it can be seen that another way to obtain the data MSE with source precoder is to use the substitutions $\mathbf{\Psi}_{sr,k} \rightarrow \mathbf{P}_k^{\text{H}}\mathbf{\Psi}_{sr,k}\mathbf{P}_k$, $\hat{\mathbf{H}}_{sr,k} \rightarrow \hat{\mathbf{H}}_{sr,k}\mathbf{P}_k$, and $\mathbf{R}_{\mathbf{s}_k} \rightarrow \mathbf{I}_{N_k}$, in (28).

With the additional power constraint for the source precoders, the optimization problem of joint transceiver design is formulated as

$$\begin{aligned}\min_{\mathbf{G}_k,\mathbf{F}_k,\mathbf{P}_k} \quad & \sum_k \text{MSE}_k(\mathbf{G}_k, \mathbf{F}_k, \mathbf{P}_k)\\ \text{s.t.} \quad & \sum_k \text{Tr}\left(\mathbf{P}_k\mathbf{P}_k^{\text{H}}\right) \le P_s\\ & \sum_k \text{Tr}\left(\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\mathbf{F}_k^{\text{H}}\right) \le P_r\end{aligned} \tag{71}$$

where P_s is the maximum transmit power at the source. In general, the optimization problem (71) is nonconvex with respective to the three design variables, and there is no closed-form solution. However, when $\mathbf{P}_k$'s are fixed, the solution for $\mathbf{G}_k$'s and $\mathbf{F}_k$'s can be directly obtained from results given by (46) and (47) with substitutions $\mathbf{\Psi}_{sr,k} \rightarrow \mathbf{P}_k^{\text{H}}\mathbf{\Psi}_{sr,k}\mathbf{P}_k$, $\hat{\mathbf{H}}_{sr,k} \rightarrow \hat{\mathbf{H}}_{sr,k}\mathbf{P}_k$, and $\mathbf{R}_{\mathbf{s}_k} \rightarrow \mathbf{I}_{N_k}$. On the other hand, when $\mathbf{G}_k$'s and $\mathbf{F}_k$'s are fixed, the optimization problem (71) is convex with respect to $\mathbf{P}_k$'s. Therefore, an iterative algorithm can be employed for joint design of source precoder, relay forwarding matrix and destination equalizer.

In order to solve $\mathbf{P}_k$'s when $\mathbf{G}_k$'s and $\mathbf{F}_k$'s are fixed, the data MSE (69) is rewritten as

$$\begin{aligned}\text{MSE}_k&(\mathbf{G}_k, \mathbf{F}_k, \mathbf{P}_k)\\ &= \text{Tr}\left(\mathbf{P}_k^{\text{H}}(\mathbf{G}_k\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\hat{\mathbf{H}}_{sr,k})^{\text{H}}(\mathbf{G}_k\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\hat{\mathbf{H}}_{sr,k})\mathbf{P}_k\right)\\ &\quad - \text{Tr}\left(\mathbf{P}_k^{\text{H}}\hat{\mathbf{H}}_{sr,k}^{\text{H}}\mathbf{F}_k^{\text{H}}\hat{\mathbf{H}}_{rd,k}^{\text{H}}\mathbf{G}_k^{\text{H}}\right)\\ &\quad - \text{Tr}\left(\mathbf{G}_k\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\hat{\mathbf{H}}_{sr,k}\mathbf{P}_k\right) + \text{Tr}\left(\mathbf{P}_k^{\text{H}}\mathbf{N}_k\mathbf{P}_k\right)\\ &\quad + \text{Tr}\left(\mathbf{I}_{N_k}\right) + a_k\end{aligned} \tag{72}$$

with

$$\begin{aligned}\mathbf{N}_k \triangleq\, & \text{Tr}\left(\mathbf{G}_k^{\text{H}}\mathbf{G}_k\right)\left(\hat{\mathbf{H}}_{sr,k}^{\text{H}}\mathbf{F}_k^{\text{H}}\mathbf{F}_k\hat{\mathbf{H}}_{sr,k}\lambda_{\max}(\mathbf{\Psi}_{rd,k})\right.\\ &\qquad \left. + \lambda_{\max}(\mathbf{\Psi}_{rd,k})\text{Tr}\left(\mathbf{F}_k^{\text{H}}\mathbf{F}_k\right)\mathbf{\Psi}_{sr,k}\right)\\ & + \text{Tr}\left(\mathbf{F}_k^{\text{H}}\hat{\mathbf{H}}_{rd,k}^{\text{H}}\mathbf{G}_k^{\text{H}}\mathbf{G}_k\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\right)\mathbf{\Psi}_{sr,k}\end{aligned} \tag{73}$$

$$\begin{aligned}a_k = & \left(\sigma_{n_2}^2 + \sigma_{n_1}^2\,\text{Tr}\left(\mathbf{F}_k\mathbf{F}_k^{\text{H}}\mathbf{\Psi}_{rd,k}\right)\right)\text{Tr}\left(\mathbf{G}_k\mathbf{G}_k^{\text{H}}\right)\\ & + \sigma_{n_1}^2\,\text{Tr}\left(\mathbf{G}_k\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\mathbf{F}_k^{\text{H}}\hat{\mathbf{H}}_{rd,k}^{\text{H}}\mathbf{G}_k^{\text{H}}\right).\end{aligned} \tag{74}$$

In (73), we have used the spectral approximation $\mathbf{\Psi}_{rd,k} \approx \lambda_{\max}(\mathbf{\Psi}_{rd,k})\mathbf{I}_{N_R}$, so that the objective function for designing $\mathbf{P}_k$'s is consistent with that of $\mathbf{F}_k$'s and $\mathbf{G}_k$'s. However, if there is no correlation in the second hop channel estimation error, $\mathbf{\Psi}_{rd,k} = \lambda_{\max}(\mathbf{\Psi}_{rd,k})\mathbf{I}_{N_R}$ and there is no approximation.

Notice that the data MSE (72) is equivalent to the following expression involving Frobenius norm

$$\begin{aligned}\text{MSE}_k&(\mathbf{G}_k, \mathbf{F}_k, \mathbf{P}_k)\\ &= \left\| \begin{bmatrix} (\mathbf{G}_k\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\hat{\mathbf{H}}_{sr,k})\mathbf{P}_k - \mathbf{I}_{N_k} \\ \mathbf{N}_k^{1/2}\mathbf{P}_k \end{bmatrix} \right\|_F^2 + a_k.\end{aligned} \tag{75}$$

Furthermore, the two power constraints in the optimization problem (71) can also be reformulated into expressions involving Frobenius norm

$$\left\| \left[\mathbf{P}_0^{\text{T}}, \cdots, \mathbf{P}_{K-1}^{\text{T}}\right]^{\text{T}} \right\|_F^2 \le P_s. \tag{76}$$

$$\begin{aligned}&\left\| \left[(\mathbf{\Gamma}_0\mathbf{P}_0)^{\text{T}}, \cdots, (\mathbf{\Gamma}_{K-1}\mathbf{P}_{K-1})^{\text{T}}\right]^{\text{T}} \right\|_F^2\\ &\qquad \le P_r - \sum_k \sigma_{n_1}^2\,\text{Tr}\left(\mathbf{F}_k\mathbf{F}_k^{\text{H}}\right)\end{aligned} \tag{77}$$

where

$$\mathbf{\Gamma}_k = \left(\text{Tr}\left(\mathbf{F}_k^{\text{H}}\mathbf{F}_k\right)\mathbf{\Psi}_{sr,k} + \hat{\mathbf{H}}_{sr,k}^{\text{H}}\mathbf{F}_k^{\text{H}}\mathbf{F}_k\hat{\mathbf{H}}_{sr,k}\right)^{1/2}. \tag{78}$$

Because the last term a_k in (72) is independent of $\mathbf{P}_k$'s, it can be neglected, and the optimization problem (71) with respective to $\mathbf{P}_k$'s can be formulated as the following second-order conic programming (SOCP) problem [see (79) at the bottom of the page]. This problem can be efficiently solved by using interpoint polynomial algorithms [28].

When $\mathbf{P}_k$'s are fixed, the proposed solutions for $\mathbf{F}_k$'s and $\mathbf{G}_k$'s in the previous section are the optimal solution for the corresponding optimization problem. On the other hand, when $\mathbf{F}_k$'s and $\mathbf{G}_k$'s are fixed, the solution for $\mathbf{P}_k$'s obtained from the SOCP problem is also the optimal solution. It means that the objective function of joint transceiver design monotonically decreases at each iteration, and the proposed iterative algorithm converges.

VII. 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 estimation 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, $N_S = M_R = N_R = M_D = 4$ is considered. The number of subcarriers K is set to be 64, and the length of the multipath channels in both hops is $L = 4$. The channel impulse response is generated according to the HIPERLAN/2 standard [10]. The signal-to-noise ratio (SNR) of the first hop is defined as $\mathrm{E}_s/\mathrm{N}_1 = P_s/(K\sigma_{n_1}^2)$, and is fixed as 30 dB. At the source, on each subcarrier, four independent data streams are transmitted, and QPSK is used as the modulation scheme. The SNR at the second hop is defined as $\mathrm{E}_r/\mathrm{N}_2 = P_r/(K\sigma_{n_2}^2)$. In the figures, MSE is referred to total simulated MSE over all subcarriers normalized by K. Each point in the following figures is an average of 10 000 realizations. In order to solve SOCP problems, the widely used optimization Matlab toolbox CVX is adopted [39].

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 in training sequence is demonstrated, and the training is white in time domain. 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 [22], [23], and therefore we have

$$\mathbf{D}\mathbf{D}^{\mathrm{H}} = \mathbf{I}_L \otimes K \begin{bmatrix} 1 & \alpha & \alpha^2 & \alpha^3 \\ \alpha & 1 & \alpha & \alpha^2 \\ \alpha^2 & \alpha & 1 & \alpha \\ \alpha^3 & \alpha^2 & \alpha & 1 \end{bmatrix}. \tag{80}$$

It is assumed that the same training sequence is used for channel estimation in the two hops. Based on the definition of $\mathbf{\Psi}_{sr,k}$ and $\mathbf{\Psi}_{rd,k}$ in (24) and (26), and together with (80), we have

$$\mathbf{\Psi}_{sr,k} = \mathbf{\Psi}_{rd,k} = \sigma_e^2 \begin{bmatrix} 1 & \alpha & \alpha^2 & \alpha^3 \\ \alpha & 1 & \alpha & \alpha^2 \\ \alpha^2 & \alpha & 1 & \alpha \\ \alpha^3 & \alpha^2 & \alpha & 1 \end{bmatrix}^{-1} \tag{81}$$

where $\sigma_e^2 = 1/\mathrm{SNR}_e$ can be viewed as the variance of channel estimation errors and SNR_e is SNR during channel estimation process.

First, we investigate the performance of the proposed algorithm with fixed source precoder $\mathbf{P}_k = \mathbf{I}_4$ and when $\alpha = 0.4$ in (81). 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 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 algorithm over the algorithm based on only estimated CSI enlarges when σ_e^2 increases.

Fig. 3 shows the MSE of the output data at the destination for both proposed algorithm and the algorithm based on estimated CSI only with fixed source precoder $\mathbf{P}_k = \mathbf{I}_4$ and with different α. It can be seen that although performance degradation is observed for both algorithms when α increases, the proposed 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 preferred in channel estimation.

$$\begin{aligned} \min_{\mathbf{P}_k, t} \quad & t \\ \text{s.t.} \quad & \left\| \begin{bmatrix} (\mathbf{G}_0\hat{\mathbf{H}}_{rd,0}\mathbf{F}_0\hat{\mathbf{H}}_{sr,0})\mathbf{P}_0 - \mathbf{I}_{N_0} \\ \mathbf{N}_0^{1/2}\mathbf{P}_0 \\ \vdots \\ (\mathbf{G}_{K-1}\hat{\mathbf{H}}_{rd,K-1}\mathbf{F}_{K-1}\hat{\mathbf{H}}_{sr,K-1})\mathbf{P}_{K-1} - \mathbf{I}_{N_{K-1}} \\ \mathbf{N}_{K-1}^{1/2}\mathbf{P}_{K-1} \end{bmatrix} \right\|_F \leq t \\ & \left\| \left[\mathbf{P}_0^{\mathrm{T}}, \cdots, \mathbf{P}_{K-1}^{\mathrm{T}}\right]^{\mathrm{T}} \right\|_F \leq \sqrt{P_s} \\ & \left\| \left[(\mathbf{\Gamma}_0\mathbf{P}_0)^{\mathrm{T}}, \cdots, (\mathbf{\Gamma}_0\mathbf{P}_0)^{\mathrm{T}}\right]^{\mathrm{T}} \right\|_F \leq \sqrt{P_r - \sum_k \sigma_{n_1}^2 \mathrm{Tr}\left(\mathbf{F}_k\mathbf{F}_k^{\mathrm{H}}\right)}. \end{aligned} \tag{79}$$

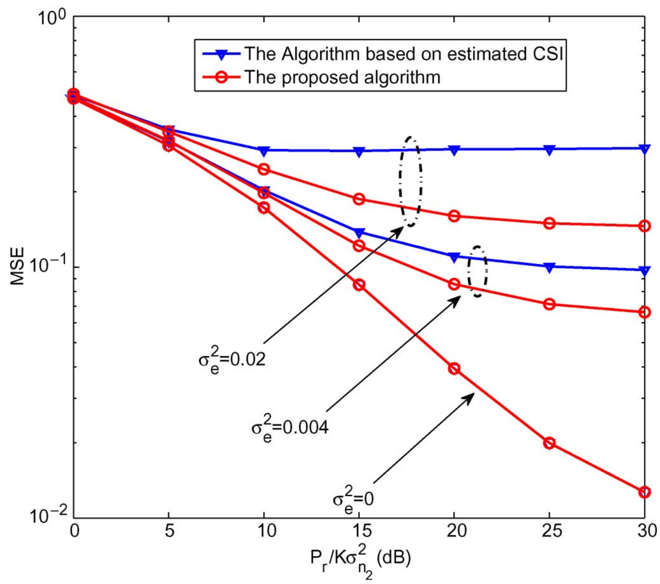


Fig. 2. MSE of received signal at the destination for different σ_e^2 when $\alpha = 0.4$ and with $\mathbf{P}_k = \mathbf{I}_4$.

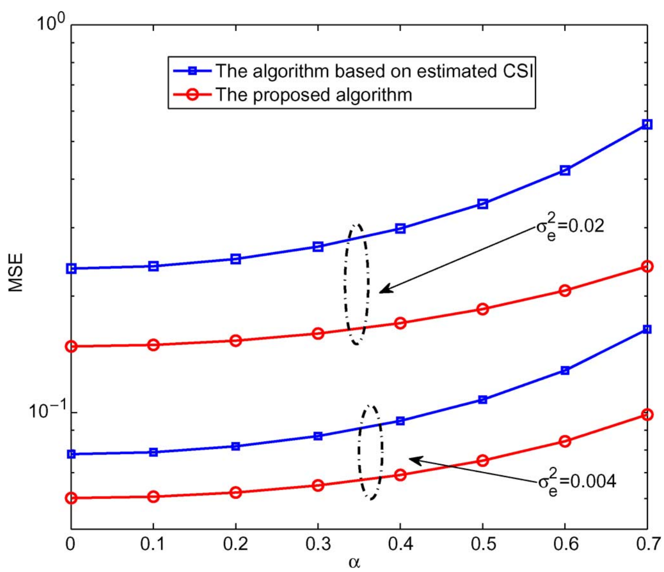


Fig. 3. MSE of received signal at the destination for different α when $P_r/K\sigma_{n_2}^2 = 30$ dB and with $\mathbf{P}_k = \mathbf{I}_4$.

Fig. 4 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.

When source precoder design is considered, the proposed algorithm is an iterative algorithm. Fig. 5 shows the convergence behavior of the proposed iterative algorithm with different initial values of $\mathbf{P}$. In the figure, the suboptimal solution as the initial value for $\mathbf{P}$ refers to the solution given in [24] based on the first hop CSI. It can be seen that the proposed algorithm with suboptimal solution as initial value has a faster convergence speed than that with identity matrix as the initial value.

Fig. 6 compares the data MSEs of the proposed iterative algorithm under channel uncertainties and the iterative algorithm based on estimated CSI only in [20]. Similar to the case

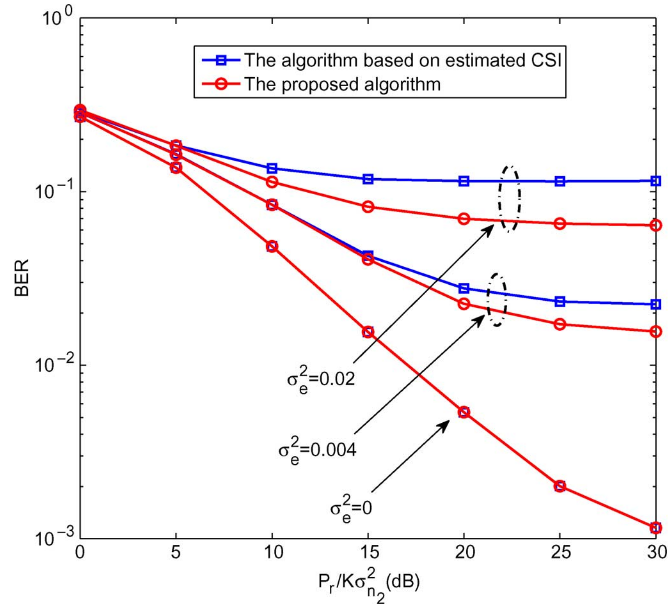


Fig. 4. BER of received data at the destination for different σ_e^2 when $\alpha = 0.5$ and with $\mathbf{P}_k = \mathbf{I}_4$.

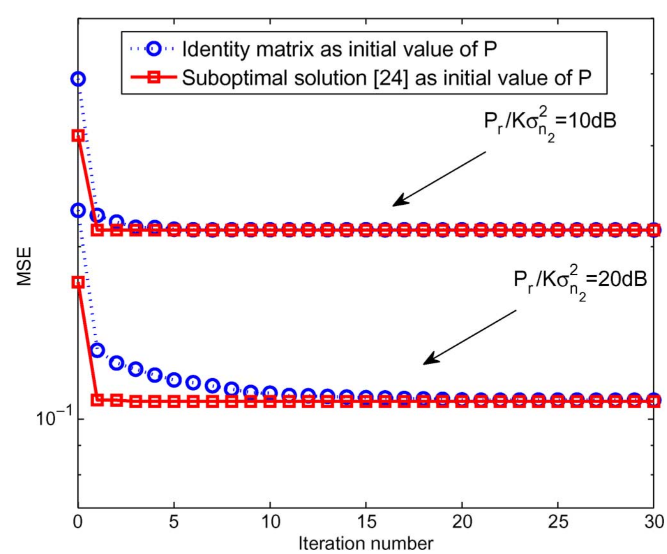


Fig. 5. Convergence behavior of the proposed iterative algorithm when $\alpha = 0.4$ and $\sigma_e^2 = 0.01$.

with fixed source precoder, the proposed joint design algorithm taking into account the channel estimation uncertainties performs better than the algorithm based on estimated CSI only.

Finally, Fig. 7 illustrates the data MSE of the iterative transceiver design algorithm based on estimated CSI only [20] and the proposed algorithms with source precoder jointly designed or simply set to $\mathbf{P}_k = \mathbf{I}_4$. It can be seen that when CSI is perfectly known ($\sigma_e^2 = 0$), the algorithms with source precoder design performs better than that by setting precoder $\mathbf{P}_k = \mathbf{I}_4$. On the other hand, when $\sigma_e^2 \geq 0.004$, even the proposed algorithm with simple precoder $\mathbf{P}_k = \mathbf{I}_4$ performs better than the algorithm based on estimated CSI only with source precoder design. Furthermore, when the channel estimation errors increases, the performance gap between the proposed algorithms with and without source precoder design decreases. Notice that

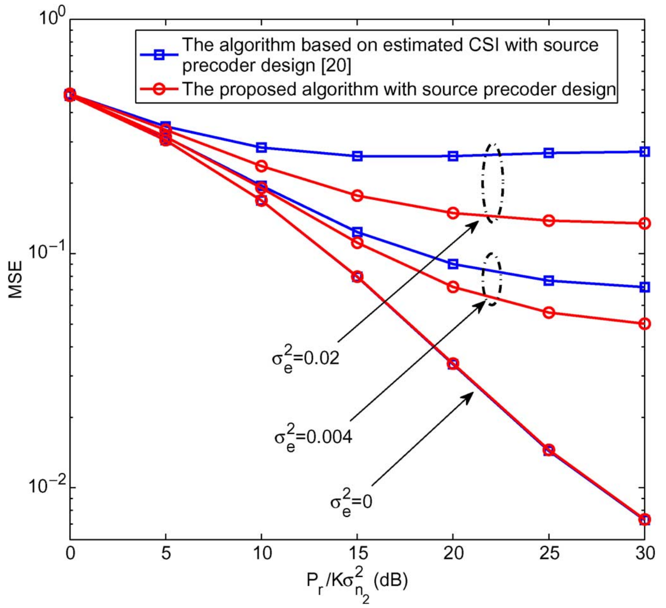


Fig. 6. MSE of received signal at the destination for different σ_e^2 when $\alpha = 0.4$.

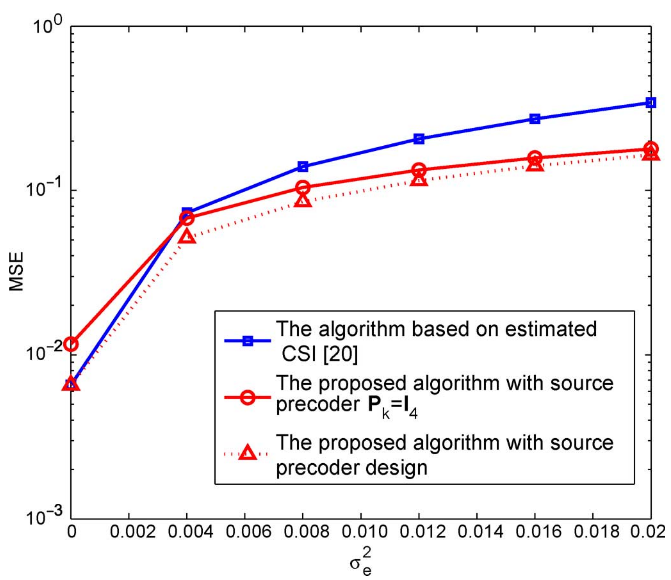


Fig. 7. MSE of received data at the destination for different σ_e^2, when $\alpha = 0.5$ and $P_r/K\sigma_{n_2}^2 = 30$ dB.

the algorithm without source precoder design has a much lower complexity, thus it represents a promising tradeoff in terms of complexity and performance.

VIII. Conclusion

In this paper, linear transceiver design was addressed for AF MIMO-OFDM relaying systems with channel estimation errors based on MMSE criterion. The linear channel estimators and the corresponding MSE expressions were first derived. Then a general solution for optimal relay forwarding matrix and destination 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. Furthermore, the design was extended to the case where source precoder design is involved. Simulation results showed that the proposed algorithms offer significant performance improvements over the algorithms 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 given by (82) at the bottom of the page, whose 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 the relationship between $\mathbf{d}$ and $\mathbf{s}$ which is given by (83).

$$\mathbf{d} = \left[\Bigg(\underbrace{\frac{1}{\sqrt{K}} \sum_{k=0}^{K-1} \mathbf{s}_k e^{j\frac{2\pi}{K}k(0)}}_{\mathbf{d}_0} \Bigg)^{\mathrm{T}} \Bigg(\underbrace{\frac{1}{\sqrt{K}} \sum_{k=0}^{K-1} \mathbf{s}_k e^{j\frac{2\pi}{K}k(1)}}_{\mathbf{d}_1} \Bigg)^{\mathrm{T}} \cdots \Bigg(\underbrace{\frac{1}{\sqrt{K}} \sum_{k=0}^{K-1} \mathbf{s}_k e^{j\frac{2\pi}{K}k(K-1)}}_{\mathbf{d}_{K-1}} \Bigg)^{\mathrm{T}} \right]^{\mathrm{T}} \quad (83)$$

From (82) and (83), by straightforward computation, the signal model given in (6) can be reformulated as

$$\begin{aligned} \mathbf{r} &= \boldsymbol{\mathcal{H}}_{sr}\mathbf{d} + \mathbf{v} \\ &= \mathrm{vec}\left(\left[\boldsymbol{\mathcal{H}}_{sr}^{(0)} \cdots \boldsymbol{\mathcal{H}}_{sr}^{(L_1-1)} \right] \mathbf{D} \right) + \mathbf{v} \\ &= \left(\mathbf{D}^{\mathrm{T}} \otimes \mathbf{I}_{M_R} \right) \mathrm{vec}\left(\left[\boldsymbol{\mathcal{H}}_{sr}^{(0)} \cdots \boldsymbol{\mathcal{H}}_{sr}^{(L_1-1)} \right] \right) + \mathbf{v} \quad (84) \end{aligned}$$

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

$$\boldsymbol{\mathcal{H}}_{sr} \triangleq \begin{bmatrix} \boldsymbol{\mathcal{H}}_{sr}^{(0)} & \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\mathcal{H}}_{sr}^{(L_1-1)} & \boldsymbol{\mathcal{H}}_{sr}^{(L_1-2)} & \cdots & \boldsymbol{\mathcal{H}}_{sr}^{(1)} \\ \boldsymbol{\mathcal{H}}_{sr}^{(1)} & \boldsymbol{\mathcal{H}}_{sr}^{(0)} & \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\mathcal{H}}_{sr}^{(L_1-1)} & \cdots & \boldsymbol{\mathcal{H}}_{sr}^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\mathcal{H}}_{sr}^{(L_1-1)} & \boldsymbol{\mathcal{H}}_{sr}^{(L_1-2)} & \boldsymbol{\mathcal{H}}_{sr}^{(L_1-3)} & \cdots & \boldsymbol{\mathcal{H}}_{sr}^{(0)} \end{bmatrix}, \quad (82)$$

APPENDIX B
PROOF OF (17)

For the expectation of the following product

$$\mathbf{\Sigma} = \mathbb{E}\{\mathbf{QRW}^{\rm H}\} \tag{85}$$

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

$$\begin{aligned}\mathbf{\Sigma}(i,j) &= \mathbb{E}\left\{\mathbf{Q}(i,:)\mathbf{R}\mathbf{W}(j,:)^{\rm H}\right\}\\ &= \sum_t\sum_k \mathbb{E}\left\{\mathbf{Q}(i,t)\mathbf{R}(t,k)\mathbf{W}(j,k)^*\right\}.\end{aligned} \tag{86}$$

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

$$\mathbb{E}\left\{\mathrm{vec}(\mathbf{Q})\mathrm{vec}^{\rm H}(\mathbf{W})\right\} = \mathbf{A}\otimes\mathbf{B} \tag{87}$$

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, (86) can be further written as

$$\begin{aligned}\mathbf{\Sigma}(i,j) &= \sum_t\sum_k \left(\mathbf{R}(t,k)\mathbb{E}\left\{\mathbf{Q}(i,t)\mathbf{W}(j,k)^*\right\}\right)\\ &= \sum_t\sum_k \mathbf{R}(t,k)\mathbf{A}(t,k)\mathbf{B}(i,j).\end{aligned} \tag{88}$$

Finally, writing (88) back to matrix form, we have [37]

$$\mathbf{\Sigma} = \mathbf{B}\mathrm{Tr}(\mathbf{R}\mathbf{A}^{\rm T}). \tag{89}$$

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 $\mathbf{A} = \sum_{\ell_2=0}^{L_1-1}\sum_{\ell_1=0}^{L_1-1}(e^{-j(2\pi/K)k(\ell_1-\ell_2)}\mathbf{\Phi}_{\ell_1,\ell_2}^{sr})$, $\mathbf{B} = \mathbf{I}_{M_R}$ and $\mathbf{Q} = \mathbf{W} = \Delta\mathbf{H}_{sr,k}$, into (89), we have (17).

APPENDIX C
PROOF OF PROPERTY 1

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

$$\begin{aligned}\mathbf{G}_k&\left(\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\mathbf{F}_k^{\rm H}\hat{\mathbf{H}}_{rd,k}^{\rm H}+\mathbf{K}_k\right)\mathbf{G}_k^{\rm H}\\ &= \mathbf{R}_{s_k}(\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\hat{\mathbf{H}}_{sr,k})^{\rm H}\mathbf{G}_k^{\rm H}.\end{aligned} \tag{90}$$

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

$$\begin{aligned}&\mathbf{F}_k^{\rm H}\hat{\mathbf{H}}_{rd,k}^{\rm H}\mathbf{G}_k^{\rm H}\mathbf{G}_k\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}+\mathbf{F}_k^{\rm H}\mathrm{Tr}\left(\mathbf{G}_k\mathbf{G}_k^{\rm H}\right)\\ &\quad\times\mathbf{\Psi}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}+\gamma_k\mathbf{F}_k^{\rm H}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\\ &= \mathbf{F}_k^{\rm H}\left(\hat{\mathbf{H}}_{sr,k}\mathbf{R}_{s_k}\mathbf{G}_k\hat{\mathbf{H}}_{rd,k}\right)^{\rm H}.\end{aligned} \tag{91}$$

After taking the traces of both sides of (90) and (91) and with the fact that the traces of their right-hand sides are equivalent, i.e.,

$$\begin{aligned}\mathrm{Tr}&\left(\mathbf{R}_{s_k}(\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\hat{\mathbf{H}}_{sr,k})^{\rm H}\mathbf{G}_k^{\rm H}\right)\\ &= \mathrm{Tr}\left(\mathbf{F}_k^{\rm H}\left(\hat{\mathbf{H}}_{sr,k}\mathbf{R}_{s_k}\mathbf{G}_k\hat{\mathbf{H}}_{rd,k}\right)^{\rm H}\right)\end{aligned}$$

we directly have

$$\begin{aligned}\mathrm{Tr}&\left(\mathbf{G}_k\left(\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\mathbf{F}_k^{\rm H}\hat{\mathbf{H}}_{rd,k}^{\rm H}+\mathbf{K}_k\right)\mathbf{G}_k^{\rm H}\right)\\ &= \mathrm{Tr}\left(\mathbf{F}_k^{\rm H}\hat{\mathbf{H}}_{rd,k}^{\rm H}\mathbf{G}_k^{\rm H}\mathbf{G}_k\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\right)\\ &\quad+\gamma_k\mathrm{Tr}\left(\mathbf{F}_k^{\rm H}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\right)\\ &\quad+\mathrm{Tr}\left(\mathbf{G}_k\mathbf{G}_k^{\rm H}\right)\mathrm{Tr}\left(\mathbf{F}_k^{\rm H}\mathbf{\Psi}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\right).\end{aligned} \tag{92}$$

By the property of trace operator

$$\begin{aligned}\mathrm{Tr}&\left(\mathbf{G}_k\left(\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\mathbf{F}_k^{\rm H}\hat{\mathbf{H}}_{rd,k}^{\rm H}\right)\mathbf{G}_k^{\rm H}\right)\\ &= \mathrm{Tr}\left(\mathbf{F}_k^{\rm H}\hat{\mathbf{H}}_{rd,k}^{\rm H}\mathbf{G}_k^{\rm H}\mathbf{G}_k\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\right)\end{aligned}$$

and (92) reduces to

$$\begin{aligned}\mathrm{Tr}\left(\mathbf{G}_k\mathbf{K}_k\mathbf{G}_k^{\rm H}\right) &= \mathrm{Tr}\left(\mathbf{G}_k\mathbf{G}_k^{\rm H}\right)\mathrm{Tr}\left(\mathbf{F}_k^{\rm H}\mathbf{\Psi}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\right)\\ &\quad+\gamma_k\mathrm{Tr}\left(\mathbf{F}_k^{\rm H}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\right).\end{aligned} \tag{93}$$

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

$$\begin{aligned}\mathrm{Tr}\left(\mathbf{G}_k\mathbf{K}_k\mathbf{G}_k^{\rm H}\right) &= \mathrm{Tr}\left(\mathbf{G}_k\mathbf{G}_k^{\rm H}\right)\mathrm{Tr}\left(\mathbf{F}_k^{\rm H}\mathbf{\Psi}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\right)\\ &\quad+\mathrm{Tr}\left(\mathbf{G}_k\mathbf{R}_{n_2,k}\mathbf{G}_k^{\rm H}\right).\end{aligned} \tag{94}$$

Comparing (93) with (94), it can be concluded that

$$\mathrm{Tr}\left(\mathbf{G}_k\mathbf{R}_{n_2,k}\mathbf{G}_k^{\rm H}\right) = \gamma_k\mathrm{Tr}\left(\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\mathbf{F}_k^{\rm H}\right). \tag{95}$$

Putting (95) into (34c), we have $\mathrm{Tr}(\mathbf{G}_k\mathbf{R}_{n_2,k}\mathbf{G}_k^{\rm H}) - \gamma_k P_{r,k} = 0$. As $\mathbf{R}_{n_2,k} = \sigma_{n_2}^2\mathbf{I}_{M_D}$, it is straightforward that

$$\sigma_{n_2}^2\mathrm{Tr}\left(\mathbf{G}_k\mathbf{G}_k^{\rm H}\right) = \gamma_k P_{r,k}. \tag{96}$$

Furthermore, based on the fact $\gamma_0 = \gamma_1 = \cdots = \gamma_{K-1} = \rho$ and taking summation of both sides of (96), the following equation holds:

$$\sum_k \sigma_{n_2}^2\mathrm{Tr}\left(\mathbf{G}_k\mathbf{G}_k^{\rm H}\right) = \rho\sum_k P_{r,k}. \tag{97}$$

Putting (97) into (34e), we have

$$\sum_k \sigma_{n_2}^2\mathrm{Tr}\left(\mathbf{G}_k\mathbf{G}_k^{\rm H}\right) - \rho P_r = 0 \tag{98}$$

and it follows that

$$\gamma_k = \rho = \sigma_{n_2}^2\frac{\sum_k\mathrm{Tr}\left(\mathbf{G}_k\mathbf{G}_k^{\rm H}\right)}{P_r}. \tag{99}$$

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

$$\mathrm{Tr}\left(\mathbf{F}_{k,\mathrm{opt}}\mathbf{R}_{\mathbf{x}_k}\mathbf{F}_{k,\mathrm{opt}}^{\mathrm{H}}\right) = P_{r,k}. \tag{100}$$

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

$$\sum_k P_{r,k} = P_r. \tag{101}$$

Finally, (96) 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

$$\mathbf{F}_k = \mathbf{M}_k^{-\frac{1}{2}}\mathbf{U}_{\mathbf{\Theta}_k}\mathbf{\Sigma}_{\mathbf{F}_k}\mathbf{U}_{\mathbf{T}_k}^{\mathrm{H}}\mathbf{R}_{\mathbf{x}_k}^{-\frac{1}{2}} \tag{102}$$

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

Putting (102) into (34a), and with the following definitions [the same as the definitions in (38) and (39)]:

$$\mathbf{M}_k^{-\frac{\mathrm{H}}{2}}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}}\hat{\mathbf{H}}_{rd,k}\mathbf{M}_k^{-\frac{1}{2}} = \mathbf{U}_{\mathbf{\Theta}_k}\mathbf{\Lambda}_{\mathbf{\Theta}_k}\mathbf{U}_{\mathbf{\Theta}_k}^{\mathrm{H}} \tag{103}$$

$$\mathbf{R}_{\mathbf{x},k}^{-\frac{1}{2}}\hat{\mathbf{H}}_{sr,k}\mathbf{R}_{s,k} = \mathbf{U}_{\mathbf{T}_k}\mathbf{\Lambda}_{\mathbf{T}_k}\mathbf{V}_{\mathbf{T}_k}^{\mathrm{H}} \tag{104}$$

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

$$\begin{aligned}\mathbf{G}_k &= \mathbf{R}_{\mathbf{s}_k}(\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\hat{\mathbf{H}}_{sr,k})^{\mathrm{H}} \\ &\quad\times\left(\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}\mathbf{F}_k^{\mathrm{H}}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}} + \eta_k\mathbf{I}_{M_D}\right)^{-1} \\ &= \left(\mathbf{R}_{\mathbf{x}_k}^{-\frac{1}{2}}\hat{\mathbf{H}}_{sr,k}\mathbf{R}_{\mathbf{s}_k}\right)^{\mathrm{H}} \\ &\quad\times\left(\mathbf{R}_{\mathbf{x}_k}^{\frac{1}{2}}\mathbf{F}_k^{\mathrm{H}}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}}\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}^{\frac{1}{2}} + \eta_k\mathbf{I}_{M_R}\right)^{-1} \\ &\quad\times\mathbf{R}_{\mathbf{x}_k}^{\frac{1}{2}}\mathbf{F}_k^{\mathrm{H}}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}} \\ &= \mathbf{V}_{\mathbf{T}_k}\underbrace{\mathbf{\Lambda}_{\mathbf{T}_k}^{\mathrm{H}}\left(\mathbf{\Sigma}_{\mathbf{F}_k}^{\mathrm{H}}\mathbf{\Lambda}_{\mathbf{\Theta}_k}\mathbf{\Sigma}_{\mathbf{F}_k} + \eta_k\mathbf{I}_{M_R}\right)^{-1}\mathbf{\Sigma}_{\mathbf{F}_k}^{\mathrm{H}}}_{\triangleq\mathbf{\Sigma}_{\mathbf{G}_k}} \\ &\quad\times\mathbf{U}_{\mathbf{\Theta}_k}^{\mathrm{H}}\mathbf{M}_k^{-\frac{\mathrm{H}}{2}}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}}\end{aligned} \tag{105}$$

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

Putting (96) from Appendix C into (34b), after multiplying both sides of (34b) with $\mathbf{M}_k^{-(1/2)}$, we have

$$\begin{aligned}&\mathbf{M}_k^{-\frac{1}{2}}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}}\mathbf{G}_k\hat{\mathbf{H}}_{rd,k}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}^{\frac{1}{2}} + \mathbf{M}_k^{\frac{1}{2}}\mathbf{F}_k\mathbf{R}_{\mathbf{x}_k}^{\frac{1}{2}}\frac{\gamma_k}{\sigma_{n_2}^2} \\ &\quad= \mathbf{M}_k^{-\frac{1}{2}}\left(\hat{\mathbf{H}}_{sr,k}\mathbf{R}_{s,k}\mathbf{G}_k\hat{\mathbf{H}}_{rd,k}\right)^{\mathrm{H}}\mathbf{R}_{\mathbf{x}_k}^{-\frac{1}{2}}.\end{aligned} \tag{106}$$

Then substituting $\mathbf{F}_k$ in (102) and $\mathbf{G}_k$ in (105) into (106), we have

$$\mathbf{\Sigma}_{\mathbf{F}} = \left(\mathbf{\Lambda}_{\mathbf{\Theta}_k}\mathbf{\Sigma}_{\mathbf{G}_k}^{\mathrm{H}}\mathbf{\Sigma}_{\mathbf{G}_k}\mathbf{\Lambda}_{\mathbf{\Theta}_k} + \frac{\gamma_k}{\sigma_{n_2}^2}\mathbf{I}_{N_R}\right)^{-1}\left(\mathbf{\Lambda}_{\mathbf{T}_k}\mathbf{\Sigma}_{\mathbf{G}_k}\mathbf{\Lambda}_{\mathbf{\Theta}_k}\right)^{\mathrm{H}}. \tag{107}$$

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 (107), it can be concluded that $\mathbf{\Sigma}_{\mathbf{F}_k}$ has the following form

$$\mathbf{\Sigma}_{\mathbf{F}_k} = \begin{bmatrix}\boldsymbol{A}_{\mathbf{F}_k} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}_{N_R\times M_R} \tag{108}$$

where $\boldsymbol{A}_{\mathbf{F}_k}$ is of dimension $q_k \times p_k$ and to be determined. Furthermore, putting (108) into the definition of $\mathbf{\Sigma}_{\mathbf{G}_k}$ in (105), we have

$$\mathbf{\Sigma}_{\mathbf{G}_k} = \begin{bmatrix}\boldsymbol{A}_{\mathbf{G}_k} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}_{N_S\times M_D} \tag{109}$$

where $\boldsymbol{A}_{\mathbf{G}_k}$ is of dimension $p_k \times q_k$, and to be determined. Substituting (108) and (109) into (102) and (105), it can be concluded that

$$\begin{aligned}\mathbf{F}_k &= \left(P_{r,k}\mathbf{\Psi}_{rd,k} + \sigma_{n_2}^2\mathbf{I}_{N_R}\right)^{-\frac{1}{2}} \\ &\quad\times\mathbf{U}_{\mathbf{\Theta}_k,q_k}\boldsymbol{A}_{\mathbf{F}_k}\mathbf{U}_{\mathbf{T}_k,p_k}^{\mathrm{H}}\mathbf{R}_{\mathbf{x}_k}^{-\frac{1}{2}},\end{aligned} \tag{110}$$

$$\mathbf{G}_k = \mathbf{V}_{\mathbf{T}_k,p_k}\boldsymbol{A}_{\mathbf{G}_k}\mathbf{U}_{\mathbf{\Theta}_k,q_k}^{\mathrm{H}}\left(P_{r,k}\mathbf{\Psi}_{rd,k} + \sigma_{n_2}^2\mathbf{I}_{N_R}\right)^{-\frac{\mathrm{H}}{2}}\hat{\mathbf{H}}_{rd,k}^{\mathrm{H}} \tag{111}$$

where

$$\boldsymbol{A}_{\mathbf{G}_k} = \bar{\mathbf{\Lambda}}_{\mathbf{T}_k}^{\mathrm{H}}\left(\boldsymbol{A}_{\mathbf{F}_k}^{\mathrm{H}}\bar{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}\boldsymbol{A}_{\mathbf{F}_k} + \eta_k\mathbf{I}_{p_k}\right)^{-1}\boldsymbol{A}_{\mathbf{F}_k}^{\mathrm{H}} \tag{112}$$

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

$$\mathrm{Tr}\left(\boldsymbol{A}_{\mathbf{G}_k}\bar{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}\boldsymbol{A}_{\mathbf{G}_k}^{\mathrm{H}}\right) = \frac{\gamma_k}{\eta_k\sigma_{n_2}^2}\mathrm{Tr}\left(\boldsymbol{A}_{\mathbf{F}_k}^{\mathrm{H}}\boldsymbol{A}_{\mathbf{F}_k}\right). \tag{113}$$

On the other hand, substituting (111) into (96) in Appendix C, we have

$$\mathrm{Tr}\left(\boldsymbol{A}_{\mathbf{G}_k}\bar{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}\boldsymbol{A}_{\mathbf{G}_k}^{\mathrm{H}}\right) = \frac{\gamma_k}{\sigma_{n_2}^2}P_{r,k}. \tag{114}$$

Comparing (113) and (114), it follows that

$$\frac{1}{\eta_k}\mathrm{Tr}\left(\boldsymbol{A}_{\mathbf{F}_k}^{\mathrm{H}}\boldsymbol{A}_{\mathbf{F}_k}\right) = P_{r,k}. \tag{115}$$

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

$$\mathrm{MSE}_k(\mathbf{F}_k, \mathbf{G}_k) = \mathrm{Tr}\left(\bar{\mathbf{\Lambda}}^2_{\mathbf{T}_k}\left(\frac{1}{\eta_k}\boldsymbol{A}^{\mathrm{H}}_{\mathbf{F}_k}\bar{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}\boldsymbol{A}_{\mathbf{F}_k} + \mathbf{I}_{p_k}\right)^{-1}\right) + \underbrace{\mathrm{Tr}\left(\mathbf{R}_{\mathbf{s}_k}\right) - \mathrm{Tr}\left(\mathbf{R}_{\mathbf{s}_k}\hat{\mathbf{H}}^{\mathrm{H}}_{sr,k}\mathbf{R}^{-1}_{\mathbf{x}_k}\hat{\mathbf{H}}_{sr,k}\mathbf{R}_{\mathbf{s}_k}\right)}_{\triangleq c_k} \quad (116)$$

where c_k is a constant part independent of $\mathbf{F}_k$. Therefore, based on (115) and (116), the optimization problem (32) becomes as

$$\begin{aligned} \min_{\boldsymbol{A}_{\mathbf{F}_k}} \quad & \sum_k \mathrm{Tr}\left(\bar{\mathbf{\Lambda}}^2_{\mathbf{T}_k}\left(\frac{1}{\eta_k}\boldsymbol{A}^{\mathrm{H}}_{\mathbf{F}_k}\bar{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}\boldsymbol{A}_{\mathbf{F}_k} + \mathbf{I}_{p_k}\right)^{-1}\right) + c_k \\ \text{s.t.} \quad & \frac{1}{\eta_k}\mathrm{Tr}\left(\boldsymbol{A}^{\mathrm{H}}_{\mathbf{F}_k}\boldsymbol{A}_{\mathbf{F}_k}\right) = P_{r,k}, \\ & \sum_k P_{r,k} = P_r. \end{aligned} \quad (117)$$

For any given $P_{r,k}$, then the optimization problem (117) can be decoupled into a collection of the following suboptimization problems:

$$\begin{aligned} \min_{\boldsymbol{A}_{\mathbf{F}_k}} \quad & \mathrm{Tr}\left(\bar{\mathbf{\Lambda}}^2_{\mathbf{T}_k}\left(\frac{1}{\eta_k}\boldsymbol{A}^{\mathrm{H}}_{\mathbf{F}_k}\bar{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}\boldsymbol{A}_{\mathbf{F}_k} + \mathbf{I}_{p_k}\right)^{-1}\right) \\ \text{s.t.} \quad & \frac{1}{\eta_k}\mathrm{Tr}\left(\boldsymbol{A}^{\mathrm{H}}_{\mathbf{F}_k}\boldsymbol{A}_{\mathbf{F}_k}\right) = P_{r,k} \end{aligned} \quad (118)$$

where the constant part c_k is neglected. For any two $M \times M$ positive semidefinite 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 $\mathbf{Z}$ [38]. Together with the fact that elements of the diagonal matrix $\tilde{\mathbf{\Lambda}}_{\mathbf{T}_k}$ are in decreasing order, the objective function of (118) is minimized, when $(\boldsymbol{A}^{\mathrm{H}}_{\mathbf{F}_k}\bar{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}\boldsymbol{A}_{\mathbf{F}_k}/\eta_k + \mathbf{I}_{N_k})$ is a diagonal matrix with the diagonal elements in decreasing order. The objective function can be rewritten as

$$\begin{aligned} &\mathrm{Tr}\left(\bar{\mathbf{\Lambda}}^2_{\mathbf{T}_k}\left(\frac{1}{\eta_k}\boldsymbol{A}^{\mathrm{H}}_{\mathbf{F}_k}\bar{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}\boldsymbol{A}_{\mathbf{F}_k} + \mathbf{I}_{N_k}\right)^{-1}\right) \\ &\quad = \mathbf{d}^{\mathrm{T}}\left(\bar{\mathbf{\Lambda}}^2_{\mathbf{T}_k}\right)\underbrace{\mathrm{d}\left(\left(\frac{1}{\eta_k}\boldsymbol{A}^{\mathrm{H}}_{\mathbf{F}_k}\bar{\mathbf{\Lambda}}_{\mathbf{\Theta}_k}\boldsymbol{A}_{\mathbf{F}_k} + \mathbf{I}_{N_k}\right)^{-1}\right)}_{\triangleq \mathbf{b}} \\ &\quad \triangleq \boldsymbol{f}(\mathbf{b}) \end{aligned} \quad (119)$$

where $\mathbf{d}(\mathbf{Z})$ 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 $\mathbf{b}$ [38, 3.H.3]. Then, based on [15, Theorem 1], the optimal $\boldsymbol{A}_{\mathbf{F}_k}$ has the following structure:

$$\boldsymbol{A}_{\mathbf{F}_k,\mathrm{opt}} = \begin{bmatrix} \mathbf{\Lambda}_{\mathbf{F}_k,\mathrm{opt}} & \mathbf{0}_{N_k,p_k-N_k} \\ \mathbf{0}_{q_k-N_k,N_k} & \mathbf{0}_{q_k-N_k,p_k-N_k} \end{bmatrix} \quad (120)$$

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

Putting (120) into the definition of $\boldsymbol{A}_{\mathbf{G}_k,\mathrm{opt}}$ in (112), the structure of the optimal $\boldsymbol{A}_{\mathbf{G}_k,\mathrm{opt}}$ is given by

$$\boldsymbol{A}_{\mathbf{G}_k,\mathrm{opt}} = \begin{bmatrix} \mathbf{\Lambda}_{\mathbf{G}_k,\mathrm{opt}} & \mathbf{0}_{N_k,q_k-N_k} \\ \mathbf{0}_{p_k-N_k,N_k} & \mathbf{0}_{p_k-N_k,q_k-N_k} \end{bmatrix} \quad (121)$$

where $\mathbf{\Lambda}_{\mathbf{G}_k,\mathrm{opt}}$ 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 Process. Mag.*, pp. 18–29, Sep. 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. Inf. 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 Process.*, vol. 55, no. 6, pp. 2953–2604, 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 Process.*, 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 Process.*, vol. 56, no. 10, pt. 2, pp. 5062–5073, Oct. 2008.

[7] B. Khoshnevis, W. Yu, and R. Adve, "Grassmannian beamforming for MIMO amplify-and-forward relaying," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1397–1407, Oct. 2008.

[8] D. Tse and P. Viswanath*, Fundamentals of Wireless Communication*. Cambridge , U.K.: Cambridge Univ. Press, 2005.

[9] S. Ma and T.-S. Ng, "Time domain signal detection based on second-order statistics for MIMO-OFDM system," *IEEE Trans. Signal Process.*, vol. 55, no. 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 Process.*, 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 Process.*, 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, U.K.: Cambridge Univ. 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. Signal Process.*, 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 Process.*, vol. 53, no. 8, pp. 2700–2712, Aug. 2005.

[17] S. Serbetli and A. Yener, "Transceiver optimization for mutiuser MIMO systems," *IEEE Trans. Signal Process.*, 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. Signal Process.*, 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 Commun. Lett.*, vol. 12, no. 7, pp. 517–519, Jul. 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 Process.*, 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. Signal Process.*, vol. 57, no. 3, pp. 1141–1150, Mar. 2009.

[23] X. Zhang, D. P. Palomar, and B. Ottersten, “Statistically robust design of linear MIMO transceiver,” *IEEE Trans. Signal Process.*, 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 Process.*, vol. 53, no. 3, pp. 1070–1085, Mar. 2005.

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

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

[27] F. A. Dietrich, P. Breun, and W. Wolfgang, “Robust Tomlinson-Harashima precoding for the wireless broadcast channel,” *IEEE Trans. Signal Process.*, vol. 55, no. 2, pp. 631–644, Feb. 2007.

[28] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[29] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar, *Convex Analysis and Optimization*. New York: Athena Scientific, 2003.

[30] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[31] 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.

[32] M. Ghogho and A. Swami, “Training design for multipath channel and frequency-offset estimation in MIMO systems,” *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3957–3965, Oct. 2006.

[33] 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.

[34] 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. Program.*, vol. 107, pp. 155–187, 2006, Springer.

[35] J. G. Andrews, A. Ghosh, and R. Muhamed, *Fundamentals of WiMAX*. Englewood Cliffs, NJ: Prentice-Hall, 2007.

[36] F. Ohrtman and K. Roeder, *Wi-Fi Handbook: Building 802.11 b Wireless Networks*. New York: McGraw-Hill, 2003.

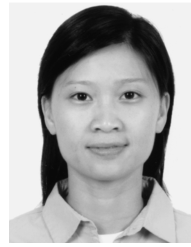
[37] S. Kay, *Fundamental of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

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

[39] M. Grant, S. Boyd, and Y. Y. Ye, CVX: Matlab Software for Disciplined Convex Programming ver. V.1.0RC3, Feb. 2007 [Online]. Available: http://www.stanford.edu/boyd/cvx/

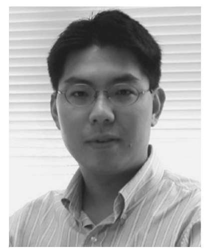
Chengwen Xing received the B.Eng. degree from Xidian University, Xi’an, China, in 2005 and the Ph.D. degree in electrical and electronic engineering from the University of Hong Kong (HKU), Hong Kong, in 2010.

Since September 2010, he has been with the School of Information and Electronics, Beijing Institute of Technology (BIT), Beijing, China, where he is currently a Lecturer. His current research interests include statistical signal processing, convex optimization, multivariate statistics, optimization, matrix analysis and cooperative communication systems.



Shaodan Ma received the B.Sc. (Eng.) and the M.Eng.Sc. degrees from NanKai University, Tianjin, China, in 1999 and 2002, respectively, all in electrical engineering, and the Ph.D. degree in electrical and electronic engineering from The University of Hong Kong (HKU), Hong Kong, in 2006.

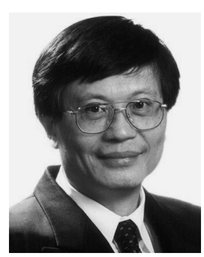
Since 2006, she has been with the Department of Electrical and Electronic Engineering, HKU, as a Postdoctoral Fellow. Her research interests include wireless communication systems, spread spectrum techniques, MIMO systems, OFDM technique, and digital signal processing.



Yik-Chung Wu received the B.Eng. (EEE) degree in 1998 and the M.Phil. degree in 2001 from The University of Hong Kong (HKU), Hong Kong. After receiving the M.S. degree, he was a Research Assistant with the same university. He received the Ph.D. degree in 2005 from Texas A&M University, College Station.

During his study at Texas A&M University, he was fully supported by the prestigious Croucher Foundation scholarship. From August 2005 to August 2006, he was with the Thomson Corporate Research, Princeton, NJ, as a Member of Technical Staff. Since September 2006, he has been with HKU as an Assistant Professor. His research interests are in general area of signal processing and communication systems, and in particular, receiver algorithm design, synchronization techniques, channel estimation, and equalization.

Dr. Wu was a TPC member for IEEE VTC Fall 2005, Globecom 2006, 2008, ICC 2007, and 2008. He is currently serving as an Associate Editor for the IEEE COMMUNICATIONS LETTERS.



Tung-Sang Ng (S’74–M’78–SM’90–F’03) received the B.Sc. (Eng.) degree from The University of Hong Kong (HKU), Hong Kong, in 1972, and the M.Eng.Sc. and Ph.D. degrees from the University of Newcastle, Australia, in 1974 and 1977, respectively, all in electrical engineering.

He worked for BHP Steel International and The University of Wollongong, Australia after graduation for 14 years before returned to HKU in 1991, where he was Professor and Chair of Electronic Engineering. He was Head of Department of Electrical and Electronic Engineering from 2000 to 2003 and Dean of Engineering from 2003 to 2007. His current research interests include wireless communication systems, spread spectrum techniques, CDMA, and digital signal processing. He has published more than 300 international journal and conference papers.

Dr. Ng was the General Chair of ISCAS’97 and the VP-Region 10 of IEEE CAS Society in 1999 and 2000. He was an Executive Committee Member and a Board Member of the IEE Informatics Divisional Board (1999–2001) and was an ordinary member of IEE Council (1999–2001). He was awarded the Honorary Doctor of Engineering Degree by the University of Newcastle in 1997, the Senior Croucher Foundation Fellowship in 1999, the IEEE Third Millenium medal in 2000, and the Outstanding Researcher Award by HKU in 2003. He is a Fellow of IET and HKIE.