| **Title** | Transceiver design for dual-hop nonregenerative MIMO-OFDM relay systems under channel uncertainties |
|-----------|--------------------------------------------------------------------------------------------------|
| **Author(s)** | Xing, C; Ma, S; Wu, YC; Ng, TS |
| **Citation** | IEEE Transactions On Signal Processing, 2010, v. 58 n. 12, p. 6325-6339 |
| **Issued Date** | 2010 |
| **URL** | http://hdl.handle.net/10722/134401 |
| **Rights** | IEEE Transactions on Signal Processing. Copyright © IEEE.; This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.; ©2010 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE. |
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-Advanced, 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 perfect 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
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 zero matrix. The notation $\text{vec}(\cdot)$ denotes the transpose, Hermitian, and conjugate of the matrix. The notation $\mathbf{1}_M$ represents Kronecker product.

transmit antennas, and one destination with $M_T$ transmit antennas, as shown in Fig. 1. At the first hop, the source transmits data to the relay, and the received signal at the relay on the $k$th subcarrier is

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

(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

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

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

(3)

(4)

where

$$\begin{align}
\mathbf{y} & \triangleq [\mathbf{y}_0^T, \ldots, \mathbf{y}_{K-1}^T]^T, & \mathbf{s} & \triangleq [\mathbf{s}_0^T, \ldots, \mathbf{s}_{K-1}^T]^T \\
\mathbf{F} & \triangleq \text{diag}[\mathbf{F}_0, \ldots, \mathbf{F}_{K-1}] \\
\mathbf{H}_{sr} & \triangleq \text{diag}[\mathbf{H}_{sr,0}, \mathbf{H}_{sr,1}, \ldots, \mathbf{H}_{sr,K-1}] \\
\mathbf{H}_{rd} & \triangleq \text{diag}[\mathbf{H}_{rd,0}, \mathbf{H}_{rd,1}, \ldots, \mathbf{H}_{rd,K-1}] \\
\mathbf{n}_1 & \triangleq [\mathbf{n}_{1,0}^T, \mathbf{n}_{1,1}^T, \ldots, \mathbf{n}_{1,K-1}^T]^T \\
\mathbf{n}_2 & \triangleq [\mathbf{n}_{2,0}^T, \mathbf{n}_{2,1}^T, \ldots, \mathbf{n}_{2,K-1}^T]^T
\end{align}$$

(5a)

(5b)

(5c)

(5d)

(5e)

(5f)

Notice that in general the matrix $\mathbf{F}$ in (4) can be an arbitrary $K N_R \times K M_R$ 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.

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

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

(1)

where $\mathbf{s}_k$ is the data vector transmitted by the source with covariance $\mathbf{R}_{s_k} = \mathbb{E} \{ \mathbf{s}_k \mathbf{s}_k^H \}$ on the $k$th subcarrier, and $\mathbf{R}_{s_k}$ 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 $\mathbf{R}_{n_{1,k}} = \sigma_{n_{1}}^2 \mathbf{I}_{M_D}$ on the $k$th subcarrier. At the relay, for each subcarrier, the received signal $\mathbf{x}_k$ is multiplied by a forwarding matrix $\mathbf{F}_k$, under a power constraint $\sum_k \text{Tr}(\mathbf{F}_k \mathbf{R}_{s_k} \mathbf{F}_k^H) \leq P_r$ where $\mathbf{R}_{s_k} = \mathbb{E} \{ \mathbf{x}_k \mathbf{x}_k^H \}$ 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

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
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

$$r = (\mathcal{F}^H \otimes I_{M_R}) x$$

$$= (\mathcal{F}^H \otimes I_{M_R}) H_{sr} (\mathcal{F} \otimes I_{N_S}) (\mathcal{F}^H \otimes I_{N_S}) s$$

$$+ (\mathcal{F}^H \otimes I_{M_R}) n_1$$

$$\triangleq d$$

(6) where $\mathcal{F}$ 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

$$r = (D^T \otimes I_{M_R}) \text{vec} \left( \left[ H_{sr}^{(0)} \cdots H_{sr}^{(L-1)} \right]^T \right) + v$$

(7) where the matrices $H_{sr}^{(0)}$ are defined as

$$H_{sr}^{(0)} = \frac{1}{K} \sum_{k=0}^{K-1} H_{sr,k} e^{j\frac{2\pi}{K} k} \quad \ell = 0, 1, \cdots, L_1 - 1.$$  

(8)

It is obvious that $H_{sr}^{(0)}$ is the $\ell$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 $D$ is a block circular matrix as

$$D = \begin{bmatrix}
    d_0 & d_1 & \cdots & \cdots & d_{K-1} \\
    d_{K-1} & d_0 & \cdots & \cdots & d_{K-2} \\
    \vdots & \vdots & \ddots & \vdots & \vdots \\
    d_{K-L_1+1} & d_{K-L_1+2} & \cdots & \cdots & d_{K-L_1}
\end{bmatrix}$$

(9) where the element $d_i$ is expressed as

$$d_i = \frac{1}{\sqrt{K}} \sum_{k=0}^{K-1} s_k e^{j\frac{2\pi}{K} k i}, \quad i = 0, \cdots, K - 1.$$  

(10)

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

$$\hat{x}_{sr} = \left( \sigma_{n_1}^{-2} (D^T \otimes I_{M_R})^H \left( D^T \otimes I_{M_R} \right) + R_{\text{channel}}^{-1} \right)^{-1} \times \sigma_{n_1}^{-2} (D^T \otimes I_{M_R})^H y$$

(11) with the corresponding MSE

$$\mathbb{E} \left\{ (\xi_{sr} - \hat{\xi}_{sr})(\xi_{sr} - \hat{\xi}_{sr})^H \right\} = \left( R_{\text{channel}}^{-1} + \sigma_{n_1}^{-2} (D^* D^T) \otimes I_{M_R} \right)^{-1}$$  

(12) where $R_{\text{channel}} = \mathbb{E} \left\{ \xi_{sr} \xi_{sr}^H \right\}$ is the prior information for channel covariance matrix. For uncorrelated channel taps, $R_{\text{channel}} = \mathcal{A}_{\text{channel}} \otimes I_{M_R N_S}$ and $\mathcal{A}_{\text{channel}} = \text{diag} \left\{ \sigma_{h_0}^2, \sigma_{h_1}^2, \cdots, \sigma_{h_{L-1}}^2 \right\}$, where $\sigma_{h_i}^2$ is the variance of the $i$th channel tap [24].

On the other hand, the channel in frequency domain and time domain has the following relationship:

$$\text{vec} \left( [H_{sr,0} \cdots H_{sr,K-1}] \right) = \sqrt{K} (F_{L_1} \otimes I_{M_R N_S}) \xi_{sr}$$  

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

$$\mathbb{E} \left\{ \text{vec} \left( [\Delta H_{sr,0} \cdots \Delta H_{sr,K-1}] \right) \right\} = \left( F_{L_1} \otimes I_{M_R N_S} \right) \left( \mathcal{A}_{\text{channel}}^{-1} \otimes I_{N_S} + \sigma_{n_1}^{-2} (D^* D^T) \right)^{-1} \otimes I_{M_R} \right) \text{vec}^H \left( [\Delta H_{sr,0} \cdots \Delta H_{sr,K-1}] \right)$$

(14) where $\Delta H_{sr,k} = H_{sr,k} - \hat{H}_{sr,k}$.

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

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

$$\mathbb{E} \left\{ \text{vec} (\Delta H_{sr,k}) \text{vec}^H (\Delta H_{sr,k}) \right\}$$

(15) where $\Phi_{0_1,0_2}^\text{sp}$ is the $N_S \times N_S$ matrix taken from the following partition of $\Phi^\text{sp}$

$$\Phi^\text{sp} = \begin{bmatrix}
    \Phi_{0,0}^\text{sp} & \Phi_{0,1}^\text{sp} & \cdots & \Phi_{0,L_1-1}^\text{sp} \\
    \vdots & \vdots & \ddots & \vdots \\
    \Phi_{L_1-1,0}^\text{sp} & \Phi_{L_1-1,1}^\text{sp} & \cdots & \Phi_{L_1-1,L_1-1}^\text{sp}
\end{bmatrix}.$$  

(16)

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

$$\mathbb{E} \left\{ \Delta H_{sr,k} R \Delta H_{sr,k}^H \right\} = \text{Tr} \left( R \sum_{l_1=0}^{L_1-1} \sum_{l_2=0}^{L_1-1} \left( e^{-j\frac{2\pi}{K} l_1} \left( \Phi_{0_1,0_2}^\text{sp} \right) \right)^T \right) I_{M_R}.$$  

(17)

\footnote{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

$$H_{rd,k} = \hat{H}_{rd,k} + \Delta H_{rd,k}, \quad k = 0, \ldots, K - 1$$  \hspace{1cm} (18)

we have the following property:

$$\text{E} \{ \Delta H_{rd,k} R \Delta H_{rd,k}^H \} = \text{Tr} \left( R \sum_{\ell_1=0}^{L_2-1} \sum_{\ell_2=0}^{L_2-1} e^{-j \frac{2\pi}{r_k} \ell_1 \Delta f} (\Phi_{\ell_1,\ell_2}^{rd})^T \right) I_{M_D}$$  \hspace{1cm} (19)

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

IV. TRANSCIEVER DESIGN PROBLEM FORMULATION

At the destination, a linear equalizer $G_k$ is adopted for each subcarrier to detect the transmitted data $s_k$ (see Fig. 1). The problem is how to design the linear forwarding matrix $F_k$ at the relay and the linear equalizer $G_k$ at the destination to minimize the MSE of the received data at the destination:

$$\text{MSE}_k(F_k, G_k) = \text{E} \left\{ \left\| (G_k H_{rd,k} F_k H_{sr,k} - I_{N_S}) s_k + G_k H_{rd,k} F_k n_{1,k} + G_k n_{2,k} \right\|^2 \right\}$$

where the expectation is taken with respect to $s_k, \Delta H_{sr,k}, \Delta H_{rd,k} n_{1,k},$ and $n_{2,k}$\(^2\). Since $s_k, n_{1,k}$, and $n_{2,k}$ are independent, the MSE expression (20) can be written as

$$\text{MSE}_k(F_k, G_k) = \text{E} \left\{ \text{Tr} \left( (G_k H_{rd,k} F_k H_{sr,k} - I_{N_S})^2 \right) \right\}$$

Applying (23) and the corresponding result for $\Delta H_{rd,k}$ to (22), the first term of $\text{MSE}_k$ becomes

$$\text{Tr} \left( (G_k E \Delta H_{rd,k} \{ H_{rd,k} F_k E \Delta H_{sr,k} \{ H_{sr,k} R_{sr} H_{sr,k}^H \} \times F_k^H H_{rd,k}^H \} G_k^H \right)$$

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

$$\Psi_{sr,k} = \sum_{\ell_1=0}^{L_2-1} \sum_{\ell_2=0}^{L_2-1} e^{-j \frac{2\pi}{r_k} \ell_1 \Delta f} (\Phi_{\ell_1,\ell_2}^{sr})^T.$$  \hspace{1cm} (24)

Similarly, the second term of $\text{MSE}_k$ in (21) can be simplified as

$$\text{Tr} \left( (G_k E \Delta H_{rd,k} \{ H_{rd,k} F_k R_{n_{1,k}} H_{sr,k}^H \} \times F_k^H H_{rd,k}^H \} G_k^H \right)$$

Based on (25) and (27), the $\text{MSE}_k$ (21) equals to

$$\text{MSE}_k(F_k, G_k) = \text{Tr} \left( G_k \left( \hat{H}_{rd,k} F_k R_{sr}, F_k^H \hat{H}_{rd,k}^H + K_k \right) G_k^H \right)$$

where $R_{sr} = \Pi_k + \sigma_{n_{2,k}}^2 I_{M_D}$

$$K_k = \left( \text{Tr} \left( F_k R_{sr} F_k^H \Psi_{sr,k} \right) + \sigma_{n_{2,k}}^2 \right) I_{M_D}$$  \hspace{1cm} (30)

Notice that the matrix $R_{sr}$ is the correlation matrix of the receive signal $x_k$ 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:

$$\min_{F_k, G_k} \sum_k \text{MSE}_k(F_k, G_k)$$

$$\text{s.t., } \sum_k \text{Tr} (F_k R_{x_k} F_k^H) \leq P_r, \quad (31)$$

**Remark 1:** In this paper, the relay estimates the source-relay channel and the destination estimates the relay-destination channel. The forwarding matrix $F_k$ and equalizer $G_k$ 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 $G_k$’S AND $F_k$’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

$$\min_{F_k, G_k, P_{r,k}} \sum_k \text{MSE}_k(F_k, G_k)$$

$$\text{s.t., } \sum_k \text{Tr} (F_k R_{x_k} F_k^H) \leq P_{r,k}, \quad k = 0, \cdots, K - 1$$

$$\sum_k P_{r,k} \leq P_r \quad (32)$$

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

$$\mathcal{L}(F_k, G_k, P_{r,k}) = \sum_k \text{MSE}_k(F_k, G_k)$$

$$+ \sum_k \gamma_k \left( \text{Tr}(F_k R_{x_k} F_k^H) - P_{r,k} \right) + \rho \left( \sum_k P_{r,k} - P_r \right) \quad (33)$$

where the positive scalars $\gamma_k$ and $\rho$ are the Lagrange multipliers. Differentiating (33) with respect to $F_k$, $G_k$ 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]

$$G_k \left( \hat{H}_{rdk} F_k R_{x_k} F_k^H \hat{H}_{rdk}^H + K_k \right) = R_{x_k} \left( \hat{H}_{rdk} F_k R_{x_k} F_k^H \hat{H}_{rdk}^H \right)^H$$

$$\hat{H}_{rdk} G_k^H G_k \hat{H}_{rdk}^H + \left( \text{Tr}(G_k^H G_k) \Psi_{rdk} + \gamma_k I_{N_{R}} \right) F_k R_{x_k}$$

$$= \left( \hat{H}_{rdk} R_{x_k} G_k \hat{H}_{rdk}^H \right)^H \quad (34a)$$

$$\gamma_k \left( \text{Tr}(F_k R_{x_k} F_k^H) - P_{r,k} \right) = 0 \quad (34c)$$

$$\gamma_k \geq 0, \quad k = 0, \cdots, K - 1 \quad (34d)$$

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

$$\gamma_0 = \gamma_1 = \cdots = \gamma_{K-1} = \rho \quad (34f)$$

$$\text{Tr} (F_k R_{x_k} F_k^H) \leq P_{r,k} \quad (34g)$$

$$\sum_k P_{r,k} \leq P_r \quad (34h)$$

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. 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 $F_k$ satisfying the KKT conditions (34a)-(34e), the power constraints (34g) and (34h) must occur on the boundaries

$$\text{Tr} (F_k R_{x_k} F_k^H) = P_{r,k} \quad (35)$$

$$\sum_k P_{r,k} = P_r \quad (36)$$

Furthermore, the corresponding $G_k$ satisfies

$$\text{Tr}(G_k^H G_k) = \gamma_k P_{r,k}/\sigma^2_{n_k}, \quad (37)$$

**Property 2:** Define the matrices $U_{T_{dk}}$, $V_{R_{dk}}$, $A_{T_{dk}}$, $U_{\Theta_k}$, and $A_{\Theta_k}$ based on eigenvalue decomposition (EVD) and singular value decomposition (SVD) as

$$(P_{r,k} \Psi_{rdk} + \sigma^2_{n_k} I_{N_{R}})^{-\frac{1}{2}} \hat{H}_{rdk}^H \hat{H}_{rdk} (P_{r,k} \Psi_{rdk} + \sigma^2_{n_k} I_{N_{R}})^{-\frac{1}{2}} = U_{\Theta_k} A_{\Theta_k} U_{\Theta_k}^H$$

$$= \hat{H}_{rdk} R_{x_k} G_k^H G_k \hat{H}_{rdk}^H \quad (38)$$

$$R_{x_k}^H \hat{H}_{rdk} R_{x_k} = U_{T_{dk}} A_{T_{dk}} V_{H_{rdk}}^H \quad (39)$$

with elements of the diagonal matrix $A_{T_{dk}}$ and $A_{\Theta_k}$ arranged in decreasing order. Then with KKT conditions (34a) and (34b), it is proved in Appendix D that the optimal forwarding matrix $F_k$ and equalizer $G_k$ must be in the form

$$F_k = \left( P_{r,k} \Psi_{rdk} + \sigma^2_{n_k} I_{N_{R}} \right)^{-\frac{1}{2}}$$

$$\times U_{\Theta_k} A_{\Theta_k} F_k U_{\Theta_k}^H \hat{H}_{rdk}^H$$

$$G_k = V_{R_{dk}} A_{R_{dk}} U_{\Theta_k}^H$$

$$\times \left( P_{r,k} \Psi_{rdk} + \sigma^2_{n_k} I_{N_{R}} \right)^{-\frac{1}{2}} \hat{H}_{rdk} \quad (40)$$

where $A_{F_k}$ and $A_{G_k}$ are to be determined. The matrix $U_{T_{dk}}$ and $V_{R_{dk}}$ are the first $p_k$ columns of $U_{T_{dk}}$ and $V_{R_{dk}}$, respectively, and $p_k = \text{Rank}(A_{T_{dk}})$. Similarly, $U_{\Theta_k}$ is the first $q_k$ columns of $U_{\Theta_k}$, and $q_k = \text{Rank}(A_{\Theta_k})$.

Right multiplying both sides of (34a) with $G_k^H$ and left multiplying both sides of (34b) with $F_k^H$ and making use of (40) and (41), the first two KKT conditions become

$$A_{G_k} \tilde{A}_{\Theta_k} A_{F_k} A_{H_{F_k}} A_{H_{G_k}} = \gamma_k G_k A_{G_k} \tilde{A}_{\Theta_k} A_{F_k} A_{H_{F_k}} \quad (42)$$

$$A_{F_k} \tilde{A}_{\Theta_k} A_{H_{F_k}} G_k A_{\Theta_k} A_{F_k} + \gamma_k \sigma_{n_k}^2 G_k A_{F_k}$$

$$= \left( A_{G_k} \tilde{A}_{\Theta_k} A_{F_k} A_{H_{F_k}} \right)^H \quad (43)$$

Notice that the solution $F_0 = \cdots = F_{K-1} = 0$ and $G_0 = \cdots = G_{K-1} = 0$ also satisfies the KKT conditions, but this solution is meaningless as no signal can be transmitted [14].
where the matrix $\mathbf{A}_{G_k}$ is the $q_k \times q_k$ principal submatrix of $\mathbf{A}_{G_k}$. Similarly, $\mathbf{A}_{T_k}$ is the $p_k \times p_k$ principal submatrix of $\mathbf{A}_{T_k}$. In this paper, we consider AF MIMO-OFDM relay systems, the matrices $\mathbf{A}_{F_k}$ and $\mathbf{A}_{G_k}$ 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 $\mathbf{A}_{F_k}$ and $\mathbf{A}_{G_k}$ have the following diagonal structure:

$$
\mathbf{A}_{F_k, \text{opt}} = \begin{bmatrix}
\mathbf{A}_{F_k, \text{opt}} & 0_{N_k, N_k} & 0_{N_k, N_k} \\
0_{N_k, N_k} & \mathbf{A}_{F_k, \text{opt}} & 0_{N_k, N_k} \\
0_{N_k, N_k} & 0_{N_k, N_k} & \mathbf{A}_{F_k, \text{opt}}
\end{bmatrix}
$$

(44)

and

$$
\mathbf{A}_{G_k, \text{opt}} = \begin{bmatrix}
\mathbf{A}_{G_k, \text{opt}} & 0_{N_k, q_k} & 0_{N_k, q_k} \\
0_{q_k, N_k} & \mathbf{A}_{G_k, \text{opt}} & 0_{q_k, N_k} \\
0_{q_k, N_k} & 0_{q_k, N_k} & \mathbf{A}_{G_k, \text{opt}}
\end{bmatrix}
$$

(45)

where $\mathbf{A}_{F_k, \text{opt}}$ and $\mathbf{A}_{G_k, \text{opt}}$ 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 $\mathbf{A}_{F_k}$ and $\mathbf{A}_{G_k}$ is unique. Now, substituting (44) and (45) into (42) and (43), and noticing that all matrices are diagonal, $\mathbf{A}_{F_k, \text{opt}}$ and $\mathbf{A}_{G_k, \text{opt}}$ can be easily solved to be

$$
\mathbf{A}_{F_k, \text{opt}} = \left(\sqrt{\frac{\sigma_{r_{2_k}}^2}{\gamma_k}} \mathbf{A}_{G_k}^{-\frac{1}{2}} \mathbf{A}_{T_k} - \eta_k \mathbf{A}_{G_k}^{-1}\right)^{+\frac{1}{2}}
$$

(46)

and

$$
\mathbf{A}_{G_k, \text{opt}} = \left(\sqrt{\frac{\sigma_{r_{2_k}}^2}{\gamma_k}} \mathbf{A}_{G_k}^{-\frac{1}{2}} \mathbf{A}_{T_k} - \frac{\gamma_k}{\sigma_{r_{2_k}}^2} \mathbf{A}_{G_k}^{-1}\right)^{+\frac{1}{2}} \mathbf{A}_{G_k}^{-\frac{1}{2}}
$$

(47)

where the matrices $\mathbf{A}_{T_k}$ and $\mathbf{A}_{G_k}$ are the principal submatrices of $\mathbf{A}_{T_k}$ and $\mathbf{A}_{G_k}$ with dimension $N_k \times N_k$, and $N_k = \min\{\text{rank}(\mathbf{A}_{T_k}), \text{rank}(\mathbf{A}_{G_k})\}$. The matrices $\mathbf{U}_{T_k, N_k}$, $\mathbf{V}_{T_k, N_k}$, and $\mathbf{U}_{E_k, N_k}$ and $\mathbf{V}_{E_k, N_k}$ are the first $N_k$ columns of $\mathbf{U}_{T_k}$, $\mathbf{V}_{T_k}$ and $\mathbf{U}_{E_k}$, $\mathbf{V}_{E_k}$, 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), $P_{T_k}$, $\gamma_k$, and $\gamma_k$ are unknown. However, notice that from (35) and (37) in Property 1, the optimal forwarding matrix and equalizer should simultaneously satisfy

$$
\text{Tr}(\mathbf{F}_{k, \text{opt}} \mathbf{R}_{z_k} \mathbf{F}_{k, \text{opt}}^H) = P_{T_k}
$$

(48)

and

$$
\text{Tr}(\mathbf{G}_{k, \text{opt}} \mathbf{G}_{k, \text{opt}}^H) = \gamma_k P_{T_k}/\sigma_{r_{2_k}}^2.
$$

(49)

Substituting (44)–(47) into (48) and (49), it can be straightforwardly shown that $\gamma_k$ and $\eta_k$ can be expressed as functions of $P_{T_k}$

$$
\gamma_k = \frac{b_{3,k} P_{T_k}}{P_{T_{1,k}} b_{1,k} + b_{1,k} b_{2,k} - b_{2,k} b_{3,k}}
$$

(50)

and

$$
\gamma_k = \frac{b_{3,k} \sigma_{r_{2_k}}^2 (P_{T_k} b_{1,k} + b_{1,k} b_{4,k} - b_{2,k} b_{3,k})}{(P_{T_k} + b_{4,k})^2 P_{T_k}}
$$

(51)

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

$$
b_{3,k} = \text{Tr} \left( \mathbf{U}_{G_k, N_k}^H (P_{T_k} \mathbf{W}_{rd_k} + \sigma_{r_{2_k}}^2 \mathbf{I}_{N_k})^{-1} \mathbf{U}_{G_k, N_k} \right)
$$

(52a)

and

$$
b_{3,k} = \frac{\gamma_k}{\sigma_{r_{2_k}}^2} \mathbf{A}_{T_k}^{-\frac{1}{2}} \mathbf{A}_{L_k}
$$

(53a)

where $\mathbf{A}_{L_k}$ 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 $\mathbf{F}_{k, \text{opt}}$ and equalizer $\mathbf{G}_{k, \text{opt}}$ are

$$
\mathbf{F}_{k, \text{opt}} = \left( P_{T_k} \mathbf{W}_{rd_k} + \sigma_{r_{2_k}}^2 \mathbf{I}_{N_k} \right)^{-\frac{1}{2}} \mathbf{U}_{G_k, N_k} \mathbf{A}_{F_k, \text{opt}}^H
$$

(54a)

and

$$
\mathbf{G}_{k, \text{opt}} = \frac{\gamma_k}{\sigma_{r_{2_k}}^2} \mathbf{A}_{T_k}^{-\frac{1}{2}} \mathbf{A}_{L_k}
$$

(54b)

with $\eta_k$ and $\gamma_k$ given by (50)–(52).

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

$$
\gamma_0 = \cdots = \gamma_{K-1}
$$

(57)

$$
\sum_k P_{T_k} = P_{T_r}
$$

(58)

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

**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., $P_{T_{1,k}} = P_{T_r}, K = 1$. 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 \((P_{r,k}\Psi_{rd,k} + \sigma^2_{n_2}I_{N_R})^{-1}\), three matrix multiplications and one EVD in (38), one matrix inversion of \(R_{xt}^{-1/2}\), 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 \(O(N_R^3)\), \(O(2N_R^3 + N_R^2M_D)\) and \(O(N_R^2)\), respectively [30]. In (39), the matrix inversion, matrix multiplications and SVD operation costs \(O(M^3)\), \(O(M^2N_S + M_R^2N_S^2)\), and \(O(M^2N_S)\), respectively. With the diagonal structures of \(\mathbf{A}_{F_{x,\text{opt}}}\) and \(\mathbf{A}_{G_{x,\text{opt}}}\), the matrix multiplications in (53) and (54) have complexities of \(O(N_R^2N_k + N_kN_R + N_kN_nM_R + M^2N_R)\) and \(O(N_RN_k + N_kN_RN_k + N_k^2N_S + N_kN_nM_D)\), respectively. On the other hand, the complexities for the two water-filling computations in (55) and (56) are \(O(N_R^2)\). 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 \(m = \text{max}\{M_D, N_R, M_R, N_S\}\).

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]:

\[
\mathbf{D}^H \propto \mathbf{I}_{N_S} \otimes \mathbf{I}_4.
\]  

Then according to (14), we have \(\Psi_{rd,k} = \sum_l \Phi_l^\text{tr} / K \propto \mathbf{I}_{N_R}\). Similarly, for the second hop, we also have

\[
\Psi_{rd,k} \propto \mathbf{I}_{N_R} \overset{\Delta}{=} \delta_{rd,k}\mathbf{I}_{N_R}
\]  

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

\[
\begin{align*}
(P_{r,k}\Psi_{rd,k} + \sigma^2_{n_2}I_{N_R})^{-\frac{1}{2}} & \hat{\mathbf{H}}^H_{rd,k} \hat{\mathbf{H}}_{rd,k} \\
\times & (P_{r,k}\Psi_{rd,k} + \sigma^2_{n_2}I_{N_R})^{-\frac{1}{2}} \\
= & \frac{1}{P_{r,k}\delta_{rd,k} + \sigma^2_{n_2}} \hat{\mathbf{H}}^H_{rd,k} \hat{\mathbf{H}}_{rd,k} \\
\end{align*}
\]  

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

\[
\mathbf{U}_{\Theta_k} = \mathbf{U}_{H_k} \mathbf{A}_{\Theta_k} = \frac{1}{(P_{r,k}\delta_{rd,k} + \sigma^2_{n_2})} \mathbf{A}_{H_k}. \tag{62}
\]

Substituting (62) into (51), \(\gamma_k\) reduces to

\[
\gamma_k = \frac{\sigma^2_{n_2}}{(P_{r,k}(1 + \delta_{rd,k}) \text{Tr}(\hat{\mathbf{A}}^{-1}_{H_k} \mathbf{A}_{H_k}^k)) + \sigma^2_{n_2} \text{Tr}(\hat{\mathbf{A}}^{-1}_{H_k} \mathbf{A}_{H_k}^k)} \tag{63}
\]

where \(\hat{\mathbf{A}}_{H_k}\) is the \(N_k \times N_k\) principal submatrix of \(\mathbf{A}_{H_k}\).

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

\[
P_{r,k} = \sqrt{\frac{\rho}{1 + \delta_{rd,k} \text{Tr}(\hat{\mathbf{A}}^{-1}_{H_k} \mathbf{A}_{H_k}^k)}} - \frac{\sigma^2_{n_2} \text{Tr}(\hat{\mathbf{A}}^{-1}_{H_k} \mathbf{A}_{H_k}^k)}{1 + \delta_{rd,k} \text{Tr}(\hat{\mathbf{A}}^{-1}_{H_k} \mathbf{A}_{H_k}^k)}, \quad k = 0, \cdots, K - 1 \tag{64}
\]

where \(\rho\) equals

\[
\rho = \sigma^2_{n_2} \left( \sum_k \frac{\text{Tr}(\hat{\mathbf{A}}^{-1}_{H_k} \mathbf{A}_{H_k}^k)}{1 + \delta_{rd,k} \text{Tr}(\hat{\mathbf{A}}^{-1}_{H_k} \mathbf{A}_{H_k}^k)} \right)^2
\]

\[
\left( P_t + \sum_k \frac{\sigma^2_{n_2} \text{Tr}(\hat{\mathbf{A}}^{-1}_{H_k} \mathbf{A}_{H_k}^k)}{1 + \delta_{rd,k} \text{Tr}(\hat{\mathbf{A}}^{-1}_{H_k} \mathbf{A}_{H_k}^k)} \right) \tag{65}
\]

B. Correlated Channel Estimation Error

Due to limited length of training sequence, \(\mathbf{D}^H \propto \mathbf{I}\) may not be possible to achieve [31]. In this case, the channel estimation errors are correlated, and \(\Psi_{rd,k} \neq \mathbf{I}\). From (38), it can be seen that the relationship between \(\mathbf{A}_{\Theta_k}\) 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)

\[
(P_{r,k}\Psi_{rd,k} + \sigma^2_{n_2}I_{N_R}) \approx (P_{r,k}\lambda_{\text{max}}(\Psi_{rd,k}) + \sigma^2_{n_2})I_{N_R}. \tag{66}
\]

For spectral approximation, \(\Psi_{rd,k}\) is replaced by \(\lambda_{\text{max}}(\Psi_{rd,k})\mathbf{I}\), where \(\lambda_{\text{max}}(\Psi_{rd,k})\) is the maximum eigenvalue of \(\Psi_{rd,k}\). 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 \(\Psi_{rd,k}\) is small, and SPA is a good approximation. With SPA, the left-hand side of (38) becomes

\[
(P_{r,k}\Psi_{rd,k} + \sigma^2_{n_2}I_{N_R})^{-\frac{1}{2}} \hat{\mathbf{H}}^H_{rd,k} \hat{\mathbf{H}}_{rd,k} \\
\times (P_{r,k}\Psi_{rd,k} + \sigma^2_{n_2}I_{N_R})^{-\frac{1}{2}} \\
\approx \frac{1}{P_{r,k}\lambda_{\text{max}}(\Psi_{rd,k}) + \sigma^2_{n_2}} \hat{\mathbf{H}}^H_{rd,k} \hat{\mathbf{H}}_{rd,k}. \tag{67}
\]
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 \( \tilde{\delta}_{rd,k} \) replaced by \( \lambda_{\text{max}}(\Psi_{rd,k}) \).

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 proceed as follows. First, with a source precoder \( \mathbf{P}_k \) before transmission, the system model in (2) is rewritten as

\[
\mathbf{y}_k = \mathbf{H}_{rd,k} \mathbf{F}_k \mathbf{H}_{sr,k} \mathbf{P}_k \mathbf{s}_k + \mathbf{H}_{rd,k} \mathbf{F}_k \mathbf{n}_{s,k} + \mathbf{n}_{2,k},
\]

(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 \( \mathbf{H}_{sr,k} \mathbf{F}_k \mathbf{P}_k \). Furthermore, without loss of generality, we can assume \( \mathbf{R}_{s,k} = \mathbf{I}_{N_s} \) in (68) as all correlations are represented by \( \mathbf{P}_k \). Then by using the substitutions \( \mathbf{H}_{sr,k} \to \mathbf{H}_{sr,k} \mathbf{P}_k \) and \( \mathbf{R}_{s,k} = \mathbf{I}_{N_s} \) 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

\[
\text{MSE}_{k}(\mathbf{G}_k, \mathbf{F}_k, \mathbf{P}_k) = \text{Tr} \left( \mathbf{G}_k \left( \mathbf{H}_{rd,k} \mathbf{F}_k \mathbf{H}_{sr,k}^H \mathbf{P}_k^H \mathbf{H}_{rd,k}^H + \mathbf{K}_k \right) \mathbf{G}_k^H \right) - \text{Tr} \left( \mathbf{P}_k \mathbf{H}_{sr,k}^H \mathbf{F}_k \mathbf{H}_{rd,k} \mathbf{G}_k \right) - \text{Tr} \left( \mathbf{G}_k \mathbf{H}_{rd,k} \mathbf{F}_k \mathbf{H}_{sr,k} \mathbf{P}_k \right) + \text{Tr} \left( \mathbf{I}_{N_s} \right)
\]

(69)

where

\[
\mathbf{R}_{s,k} = \text{Tr} \left( \mathbf{P}_k \mathbf{P}_k^H \mathbf{I}_{M_R} \right) \mathbf{I}_{M_R} + \mathbf{H}_{sr,k} \mathbf{P}_k \mathbf{P}_k^H \mathbf{H}_{sr,k}^H + \sigma^2_{n_1} \mathbf{I}_{M_R}.
\]

(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 \( \mathbf{H}_{sr,k} \to \mathbf{H}_{sr,k} \mathbf{P}_k \), \( \mathbf{H}_{sr,k} \to \mathbf{H}_{sr,k} \mathbf{P}_k \), and \( \mathbf{R}_{s,k} = \mathbf{I}_{N_s} \), in (28).

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

\[
\min_{\mathbf{G}_k, \mathbf{F}_k, \mathbf{P}_k} \sum_k \text{MSE}_k(\mathbf{G}_k, \mathbf{F}_k, \mathbf{P}_k)
\]

\[
\text{s.t.} \quad \sum_k \text{Tr} \left( \mathbf{P}_k \mathbf{P}_k^H \right) \leq P_s
\]

\[
\sum_k \text{Tr} \left( \mathbf{F}_k \mathbf{R}_{s,k} \mathbf{F}_k^H \right) \leq P_r
\]

(71)

where \( P_s \) is the maximum transmit power at the source. In general, the optimization problem (71) is nonconvex with respect 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 \( \Psi_{sr,k} \to \mathbf{P}_k \mathbf{H}_{sr,k} \mathbf{P}_k \), \( \mathbf{H}_{sr,k} \to \mathbf{H}_{sr,k} \mathbf{P}_k \), and \( \mathbf{R}_{s,k} = \mathbf{I}_{N_s} \). 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

\[
\text{MSE}_k(\mathbf{G}_k, \mathbf{F}_k, \mathbf{P}_k) = \text{Tr} \left( \mathbf{P}_k^H (\mathbf{G}_k \mathbf{H}_{rd,k} \mathbf{F}_k \mathbf{H}_{sr,k} \mathbf{H}_{rd,k}^H \mathbf{G}_k^H) \right) - \text{Tr} \left( \mathbf{P}_k^H \mathbf{H}_{sr,k}^H \mathbf{H}_{rd,k} \mathbf{G}_k^H \right) - \text{Tr} \left( \mathbf{G}_k \mathbf{H}_{rd,k} \mathbf{F}_k \mathbf{H}_{sr,k} \mathbf{P}_k \right) + \text{Tr} \left( \mathbf{P}_k^H \mathbf{N}_k \mathbf{P}_k \right) + \text{Tr} (\mathbf{I}_{N_s}) + a_k
\]

(72)

with

\[
\mathbf{N}_k = \text{Tr} \left( \mathbf{G}_k^H \mathbf{G}_k \right) \left( \mathbf{H}_{sr,k}^H \mathbf{F}_k^H \mathbf{F}_k \mathbf{H}_{sr,k} \mathbf{N}_k \lambda_{\text{max}}(\Psi_{rd,k}) \right) + \lambda_{\text{max}}(\Psi_{rd,k}) \text{Tr} \left( \mathbf{F}_k^H \mathbf{F}_k \right) \Psi_{sr,k} + \text{Tr} \left( \mathbf{F}_k^H \mathbf{F}_k \Psi_{sr,k} \right) (73)
\]

\[
\mathbf{a}_k = \left( \sigma^2_{n_2} + \sigma^2_{n_1} \text{Tr} (\mathbf{F}_k \mathbf{P}_k^H \mathbf{I}_{N_s}) \right) (\mathbf{G}_k \mathbf{G}_k^H)
\]

(74)

In (73), we have used the spectral approximation \( \Psi_{rd,k} \approx \lambda_{\text{max}}(\Psi_{rd,k}) \mathbf{I}_{N_s} \), 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, \( \Psi_{rd,k} = \lambda_{\text{max}}(\Psi_{rd,k}) \mathbf{I}_{N_s} \) and there is no approximation.

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

\[
\text{MSE}_k(\mathbf{G}_k, \mathbf{F}_k, \mathbf{P}_k) = \left\| \left[ (\mathbf{G}_k \mathbf{H}_{rd,k} \mathbf{F}_k \mathbf{H}_{sr,k} \mathbf{P}_k - \mathbf{I}_{N_s}) \right]^2 \right\|_F^2 + a_k.
\]

(75)

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

\[
\left\| \left[ \mathbf{P}_{0}^T, \cdots, \mathbf{P}_{K-1}^T \right]^T \right\|_F^2 \leq P_s.
\]

(76)

\[
\left\| \left[ (\mathbf{T}_0 \mathbf{P}_0)^T, \cdots, (\mathbf{T}_{K-1} \mathbf{P}_{K-1})^T \right]^T \right\|_F^2 \leq P_r - \sum_k \sigma^2_{n_1} \text{Tr} (\mathbf{F}_k \mathbf{F}_k^H)
\]

(77)

where

\[
\mathbf{G}_k = \left( \text{Tr} \left( \mathbf{F}_k^H \mathbf{F}_k \right) \Psi_{sr,k} + \mathbf{H}_{sr,k} \mathbf{F}_k \mathbf{H}_{sr,k} \right)^{1/2}.
\]

(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 respect to \( \mathbf{P}_k \)'s can be formulated as the following second-order conic programming (SOCP) problem [see (79) at the bottom of the
When \(P_k\)'s are fixed, the proposed solutions for \(F_k\)'s and \(G_k\)'s in the previous section are the optimal solution for the corresponding optimization problem. On the other hand, when \(F_k\)'s and \(G_k\)'s are fixed, the solution for \(P_k\)'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 \(E_b/N_0 = P_s/(K\sigma_n^2)\), 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 \(E_b/N_2 = P_r/(K\sigma_n^2)\). 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 \(D\) in (9), \(DD^H\) 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, \(DD^H\) is a block diagonal matrix, and can be written as \(DD^H = I_L \otimes \sum_\alpha d_\alpha d_\alpha^H, \) where \(\sum_\alpha d_\alpha d_\alpha^H/K\) 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

\[
DD^H = I_L \otimes K \begin{bmatrix}
1 & \alpha & \alpha^2 & \alpha^3 \\
\alpha^2 & 1 & \alpha & \alpha^2 \\
\alpha^3 & \alpha^2 & 1 & \alpha \\
\alpha^3 & \alpha^2 & 1 & 1 \\
\end{bmatrix}
\]  

(80)

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

\[
\Psi_{sr,k} = \Psi_{rd,k} = \sigma_e^2 \begin{bmatrix}
1 & \alpha & \alpha^2 & \alpha^3 \\
\alpha & 1 & \alpha & \alpha^2 \\
\alpha^2 & \alpha & 1 & \alpha \\
\alpha^3 & \alpha^2 & \alpha & 1 \\
\end{bmatrix}^{-1}
\]  

(81)

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

First, we investigate the performance of the proposed algorithm with fixed source precoder \(P_k = I_4\) and when \(\alpha = 0A\) in (81). Fig. 2 shows the MSE of the received signal at the destination with different \(\sigma_e^2\). 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 \(\sigma_e^2\) is not zero. Furthermore, the performance improvement of the proposed algorithm over the algorithm based on only estimated CSI enlarges when \(\sigma_e^2\) 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 \(P_k = I_4\) and with different \(\alpha\). It can be seen that although performance degradation is observed for both algorithms when \(\alpha\) increases, the proposed algorithm shows a significant improvement over the algorithm based on estimated CSI only. Furthermore, as \(\alpha \approx 0\) gives the best data MSE performance, it demonstrates that white sequence is preferred in channel estimation.
Fig. 2. MSE of received signal at the destination for different $\sigma_e^2$ when $\alpha = 0.4$ and with $P_k = I_4$.

Fig. 3. MSE of received signal at the destination for different $\alpha$ when $P_t/K\sigma_e^2 = 30$ dB and with $P_k = I_4$.

Fig. 4. BER of received data at the destination for different $\sigma_e^2$ when $\alpha = 0.5$ and with $P_k = I_4$.

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

Fig. 6 shows the bit error rates (BER) of the output data at the destination for different $\sigma_e^2$, 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 $P$. In the figure, the suboptimal solution as the initial value for $P$ 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 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 $P_k = I_4$. 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 $P_k = I_4$. On the other hand, when $\sigma_e^2 \geq 0.001$, even the proposed algorithm with simple precoder $P_k = I_4$ 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
When tap of the multi-path MIMO in (6), we have 
(83), which is given by (83).

\[ \mathbf{H}_{sr} = \left[ \begin{array}{cccc}
H_{sr}^{(0)} & 0 & 0 & \ldots \\
0 & H_{sr}^{(L_1-1)} & \ldots & \ldots \\
\vdots & \vdots & \ddots & \vdots \\
0 & \ldots & 0 & H_{sr}^{(L_1)}
\end{array} \right] 
\]

is the \( M \times M \) block circulant matrix given by (9). The matrix \( \mathbf{H}_{sr} \) is defined in (9).

Fig. 6. MSE of received signal at the destination for different \( \sigma_s^2 \) when \( \alpha = 0.4 \).

Fig. 7. MSE of received data at the destination for different \( \sigma_s^2 \) when \( \alpha = 0.5 \) and \( P_r/K\sigma_n^2 = 30 \) dB.

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 \( \mathbf{H}_{sr} \), defined in (6) is a \( K M_R \times K M_S \) block circulant matrix given by (82) at the bottom of the page, whose element \( H_{sr}^{(0)} \) is defined in (8). It is obvious that \( H_{sr}^{(0)} \) is the \( k \)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).

\[
\mathbf{d} = \begin{bmatrix}
\frac{1}{\sqrt{K}} \sum_{k=0}^{K-1} s_k e^{j \frac{2\pi}{K} k(0)} \\
\frac{1}{\sqrt{K}} \sum_{k=0}^{K-1} s_k e^{j \frac{2\pi}{K} k(1)} \\
\vdots \\
\frac{1}{\sqrt{K}} \sum_{k=0}^{K-1} s_k e^{j \frac{2\pi}{K} k(K-1)}
\end{bmatrix}^T + \mathbf{v} 
\]

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

\[
\mathbf{r} = \mathbf{H}_{sr} \mathbf{d} + \mathbf{v} = \mathbf{v}_c \left[ \begin{array}{c}
H_{sr}^{(0)} \\
H_{sr}^{(L_1-1)} \\
\vdots \\
H_{sr}^{(0)}
\end{array} \right] + \mathbf{v} 
\]

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

\[
\mathbf{H}_{sr} = \begin{bmatrix}
H_{sr}^{(0)} & 0 & 0 & \ldots & H_{sr}^{(L_1-1)} \\
H_{sr}^{(1)} & H_{sr}^{(0)} & 0 & \ldots & H_{sr}^{(L_1-2)} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \ldots & 0 & H_{sr}^{(L_1-1)} & H_{sr}^{(0)}
\end{bmatrix}, \tag{82}
\]
For the expectation of the following product
\[ \mathbf{\Sigma} = E\{\mathbf{QRW}^H\} \quad (85) \]
where \(\mathbf{Q}\) and \(\mathbf{W}\) are two \(M \times N\) random matrices with compatible dimension to \(\mathbf{R}\), the \((i,j)\)th element of \(\mathbf{\Sigma}\) is
\[ \mathbf{\Sigma}(i,j) = E\{\mathbf{Q}(i,:)^H \mathbf{R}(j,:) \mathbf{W}(j,:)^H\} \]
\[ = \sum_t \sum_k E\{\mathbf{Q}(i,t)\mathbf{R}(t,k)\mathbf{W}(j,k)^*\}. \quad (86) \]

If the two random matrices \(\mathbf{Q}\) and \(\mathbf{W}\) satisfy
\[ E\{\mathbf{Q}\mathbf{Q}^H\} = \mathbf{A} \otimes \mathbf{B} \quad (87) \]
where \(\mathbf{A}\) is a \(N \times N\) matrix while \(\mathbf{B}\) is a \(M \times M\) matrix, then we have the equality \(E\{\mathbf{Q}(i_1,:)^H \mathbf{W}(i_2,:)\} = \mathbf{B}(i_1,i_2)\mathbf{A}(j_1,j_2)\). As \(\mathbf{Q}(i,:)^H\) and \(\mathbf{W}(j,:)^H\) are scalars, (86) can be further written as
\[ \mathbf{\Sigma}(i,j) = \sum_t \sum_k (\mathbf{R}(t,k)E\{\mathbf{Q}(i,t)\mathbf{W}(j,k)^*\}) \]
\[ = \sum_t \sum_k \mathbf{R}(t,k)\mathbf{A}(i,j)^*\mathbf{B}(i,j). \quad (88) \]

Finally, writing (88) back to matrix form, we have [37]
\[ \mathbf{\Sigma} = \mathbf{BTr}\{\mathbf{RA}^T\}. \quad (89) \]

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 \(\mathbf{A} = \sum_{\ell_2=0}^{L_2-1} \sum_{\ell_1=0}^{L_1-1} (c_{\ell_1\ell_2})^2 \mathbf{P}^\ell_\ell_2\) 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 \(\mathbf{G}_k^H\), the following equality holds
\[ \mathbf{G}_k \left( \mathbf{H}_{rld,k} \mathbf{F}_k \mathbf{R}_{x_k} \mathbf{F}_k^H \mathbf{H}_{rld,k}^H + \mathbf{K}_k \right) \mathbf{G}_k^H \]
\[ = \mathbf{R}_{x_k} \left( \mathbf{H}_{rld,k} \mathbf{F}_k \mathbf{H}_{sr,k} \right)^H \mathbf{G}_k^H. \quad (90) \]

Left multiplying (34b) with \(\mathbf{F}_k^H\), we have
\[ \mathbf{F}_k^H \mathbf{H}_{rld,k} \mathbf{G}_k \mathbf{G}_k \mathbf{H}_{rld,k}^H + \mathbf{F}_k^H \mathbf{Tr}\{\mathbf{G}_k \mathbf{G}_k^H\} \mathbf{F}_k \mathbf{R}_{x_k} + \mathbf{F}_k^H \mathbf{Tr}\{\mathbf{G}_k \mathbf{G}_k^H\} \mathbf{F}_k \mathbf{R}_{x_k} \]
\[ = \mathbf{F}_k^H \left( \mathbf{H}_{sr,k} \mathbf{R}_{x_k} \mathbf{G}_k \mathbf{H}_{rd,k} \right)^H. \quad (91) \]

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.,
\[ \mathbf{Tr}\{\mathbf{R}_{x_k} \mathbf{H}_{sr,k} \mathbf{F}_k \mathbf{G}_k \mathbf{H}_{rd,k}^H \mathbf{G}_k^H\} = \mathbf{Tr}\left( \mathbf{F}_k^H \left( \mathbf{H}_{sr,k} \mathbf{R}_{x_k} \mathbf{G}_k \mathbf{H}_{rd,k} \right)^H \mathbf{G}_k^H\right) \]
we directly have
\[ \mathbf{Tr}\left( \mathbf{G}_k \left( \mathbf{H}_{rld,k} \mathbf{F}_k \mathbf{R}_{x_k} \mathbf{F}_k^H \mathbf{H}_{rld,k}^H + \mathbf{K}_k \right) \mathbf{G}_k^H\right) = \mathbf{Tr}\left( \mathbf{F}_k^H \mathbf{H}_{rld,k} \mathbf{G}_k \mathbf{H}_{rld,k}^H \mathbf{F}_k \mathbf{R}_{x_k} \mathbf{F}_k^H \mathbf{H}_{rld,k}^H \mathbf{F}_k \mathbf{R}_{x_k} \mathbf{F}_k^H \mathbf{G}_k^H\right) \]
\[ + \gamma_k \mathbf{Tr}\left( \mathbf{F}_k^H \mathbf{R}_{x_k}\right) + \mathbf{Tr}\left( \mathbf{G}_k \mathbf{K}_k \right) \mathbf{Tr}\left( \mathbf{F}_k^H \mathbf{R}_{x_k}\right). \quad (92) \]

By the property of trace operator
\[ \mathbf{Tr}\left( \mathbf{G}_k \left( \mathbf{H}_{rld,k} \mathbf{F}_k \mathbf{R}_{x_k} \mathbf{F}_k^H \mathbf{H}_{rld,k}^H + \mathbf{K}_k \right) \mathbf{G}_k^H\right) \]
\[ = \mathbf{Tr}\left( \mathbf{F}_k^H \mathbf{H}_{rld,k} \mathbf{G}_k \mathbf{H}_{rld,k}^H \mathbf{F}_k \mathbf{R}_{x_k}\right) \]
and (92) reduces to
\[ \mathbf{Tr}\left( \mathbf{G}_k \mathbf{K}_k \right) \mathbf{G}_k^H = \mathbf{Tr}\left( \mathbf{G}_k \mathbf{G}_k^H\right) \mathbf{Tr}\left( \mathbf{F}_k^H \mathbf{R}_{x_k}\right) \]
\[ + \gamma_k \mathbf{Tr}\left( \mathbf{F}_k^H \mathbf{R}_{x_k}\right). \quad (93) \]

On the other hand, based on the definition of \(\mathbf{K}_k\) in (30), \(\mathbf{Tr}\left( \mathbf{G}_k \mathbf{K}_k \mathbf{G}_k^H\right)\) can be also expressed as
\[ \mathbf{Tr}\left( \mathbf{G}_k \mathbf{K}_k \mathbf{G}_k^H\right) = \mathbf{Tr}\left( \mathbf{G}_k \mathbf{G}_k^H\right) \mathbf{Tr}\left( \mathbf{F}_k^H \mathbf{R}_{x_k}\right) \]
\[ + \mathbf{Tr}\left( \mathbf{G}_k \mathbf{R}_{n2,k} \mathbf{G}_k^H\right). \quad (94) \]

Comparing (93) with (94), it can be concluded that
\[ \mathbf{Tr}\left( \mathbf{G}_k \mathbf{R}_{n2,k} \mathbf{G}_k^H\right) = \gamma_k \mathbf{Tr}\left( \mathbf{F}_k \mathbf{R}_{x_k} \mathbf{F}_k^H \right). \quad (95) \]

Putting (95) into (34c), we have \(\mathbf{Tr}\left( \mathbf{G}_k \mathbf{R}_{n2,k} \mathbf{G}_k^H\right) - \gamma_k P_{r,k} = 0\). As \(\mathbf{R}_{n2,k} = \sigma_{n2}^2 \mathbf{I}_{M_D}\), it is straightforward that
\[ \sigma_{n2}^2 \mathbf{Tr}\left( \mathbf{G}_k \mathbf{G}_k^H\right) = \gamma_k P_{r,k}. \quad (96) \]

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:
\[ \sum_k \sigma_{n2}^2 \mathbf{Tr}\left( \mathbf{G}_k \mathbf{G}_k^H\right) = \rho \sum_k P_{r,k}. \quad (97) \]

Putting (97) into (34e), we have
\[ \sum_k \sigma_{n2}^2 \mathbf{Tr}\left( \mathbf{G}_k \mathbf{G}_k^H\right) - \rho P_r = 0 \]
and it follows that
\[ \rho = \sigma_{n2}^2 \sum_k \mathbf{Tr}\left( \mathbf{G}_k \mathbf{G}_k^H\right) P_r. \quad (99) \]
Since for the optimal equalizer $G_k$, $\sum_k \text{Tr}(G_k \cdot \text{opt}^H G_k \cdot \text{opt}^H) \neq 0$, it can be concluded that $\gamma_k \neq 0$. In order to have (34c) satisfied, we must have

$$\text{Tr} \left( F_k \cdot \text{opt}^H R_{x_k} \cdot F_k \cdot \text{opt}^H \right) = P_{r,k}. \quad (100)$$

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

$$\sum_k P_{r,k} = P_r. \quad (101)$$

Finally, (96) constitutes the second part of the Property 1.

### APPENDIX D

#### PROOF OF PROPERTY 2

Defining a full rank Hermitian matrix $M_k = P_{r,k} \Psi_{rd,k} + \sigma_{r2}^2 I_{NR}$, then for an arbitrary $N_R \times N_R$ matrix $F_k$, it can be written as

$$F_k = M_{k}^{-\frac{1}{2}} \Theta_k \Sigma_{F_k} U_{\Theta_k}^H R_{x_k}^{-\frac{1}{2}} \quad (102)$$

where the inner matrix $\Sigma_{F_k}$ equals $\Sigma_{F_k} = U_{\Theta_k} \Theta_k \Sigma_{F_k} U_{\Theta_k}^H$.

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

$$M_{k}^{-\frac{1}{2}} H_{rd,k}^H M_{k}^{-\frac{1}{2}} = U_{\Theta_k} \Theta_k \Sigma_{F_k} U_{\Theta_k}^H \quad (103)$$

$$R_{x_k}^{-\frac{1}{2}} H_{sr,k} R_{x_k}^{-\frac{1}{2}} = U_{\Theta_k} \Theta_k \Sigma_{F_k} U_{\Theta_k}^H \quad (104)$$

the equalizer $G_k$ can be reformulated as

$$G_k = R_{x_k} \left( H_{rd,k} F_k \cdot \Theta_k \right) \cdot \left( H_{rd,k}^H M_{k}^{-\frac{1}{2}} + \eta_k I_{MD} \right)^{-1} \times \left( H_{rd,k} \cdot \Theta_k \right) \cdot U_{\Theta_k}^H \cdot \Theta_k \Sigma_{F_k} \cdot U_{\Theta_k}^H \quad (105)$$

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 $M_{k}^{-\frac{1}{2}}$, we have

$$M_{k}^{-\frac{1}{2}} H_{rd,k}^H G_k \cdot H_{rd,k} \cdot F_k \cdot R_{x_k}^{-\frac{1}{2}} + M_{k}^{-\frac{1}{2}} F_k R_{x_k}^{-\frac{1}{2}} \cdot \frac{\gamma_k}{\sigma_{r2}^2} \quad (106)$$

Then substituting $F_k$ in (102) and $G_k$ in (105) into (106), we have

$$\Sigma_F = \left( \Theta_k \Sigma_{G_k} \cdot \Theta_k \Sigma_{G_k} \right)^{-1} \left( \Theta_k \Sigma_{G_k} \cdot \Theta_k \Sigma_{G_k} \right)^H \quad (107)$$

Since $\Theta_k$ and $\Theta_k$ are rectangular diagonal matrices (denoting their ranks by $p_k$ and $q_k$ respectively), based on (107), it can be concluded that $\Sigma_{F_k}$ has the following form

$$\Sigma_{F_k} = \left[ \begin{array}{cc} A_{F_k} & 0 \\ 0 & 0 \end{array} \right]_{N_R \times N_R} \quad (108)$$

where $A_{F_k}$ is of dimension $q_k \times p_k$ and to be determined. Furthermore, putting (108) into the definition of $\Sigma_{G_k}$ in (105), we have

$$\Sigma_{G_k} = \left[ \begin{array}{cc} A_{G_k} & 0 \\ 0 & 0 \end{array} \right]_{N_S \times N_D} \quad (109)$$

where $A_{G_k}$ 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

$$F_k = \left( P_{r,k} \Psi_{rd,k} + \sigma_{r2}^2 I_{NR} \right)^{-\frac{1}{2}} \times \left( \Theta_k \Theta_k \Sigma_{F_k} \cdot U_{\Theta_k}^H \cdot \Theta_k \Sigma_{F_k} \cdot U_{\Theta_k}^H \right) \cdot \left( H_{rd,k} \cdot \Theta_k \right) \cdot U_{\Theta_k}^H \cdot \Theta_k \Sigma_{F_k} \cdot U_{\Theta_k}^H \quad (110)$$

$$G_k = V_{r,k} A_{G_k} U_{r,k}^H \left( P_{r,k} \Psi_{rd,k} + \sigma_{r2}^2 I_{NR} \right)^{-\frac{1}{2}} H_{rd,k} \quad (111)$$

where

$$A_{G_k} = \bar{A}_{T_k} \left( A_{F_k}^H \cdot \Theta_k \Sigma_{F_k} \cdot \Theta_k \Sigma_{F_k} \cdot U_{\Theta_k}^H \right) \cdot \left( \Theta_k \Sigma_{F_k} \cdot \Theta_k \Sigma_{F_k} \cdot U_{\Theta_k}^H \right)^H \quad (112)$$

and $\bar{A}_{T_k}$ is the $p_k \times p_k$ principal submatrix of $A_{T_k}$.

### 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

$$\text{Tr} \left( A_{G_k} \cdot \Theta_k \Sigma_{G_k}^H \right) = \frac{\gamma_k}{\eta_k} \text{Tr} \left( A_{F_k}^H \cdot A_{F_k} \right). \quad (113)$$

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

$$\text{Tr} \left( A_{G_k} \cdot \bar{A}_{T_k} \cdot A_{G_k}^H \right) = \frac{\gamma_k}{\eta_k} P_{r,k}. \quad (114)$$

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

$$\frac{1}{\eta_k} \text{Tr} \left( A_{F_k}^H \cdot A_{F_k} \right) = P_{r,k}. \quad (115)$$
For the objective function in the optimization problem (32), substituting (40) and (41) into the MSE expression in (28), the MSE on the kth subcarrier can be written as

\[
\text{MSE}_k(F_k, G_k) = \text{Tr} \left( \Lambda_{T_k}^2 \left( \frac{1}{\eta_k} \Lambda_{H_k}^H \Lambda_{E_k} \Lambda_{F_k} + I_{p_k} \right)^{-1} \right) + \text{Tr} (R_{s_k} - \text{Tr} (R_{s_k} \Lambda_{H_{s_k}}^H R_{s_k}^{-1} \Lambda_{H_{s_k}} R_{s_k}))
\]

\[
\overset{\Delta}{=} c_k
\]

(116)

where \(c_k\) is a constant part independent of \(F_k\). Therefore, based on (115) and (116), the optimization problem (32) becomes as

\[
\min_{A_{F_k}} \sum_k \text{Tr} \left( \Lambda_{T_k}^2 \left( \frac{1}{\eta_k} \Lambda_{H_k}^H \Lambda_{E_k} \Lambda_{F_k} + I_{p_k} \right)^{-1} \right) + c_k
\]

s.t.

\[
\frac{1}{\eta_k} \text{Tr} (A_{H_k}^H A_{F_k}) = P_{T,k},
\]

\[
\sum_k P_{T,k} = P_T
\]

(117)

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

\[
\min_{A_{F_k}} \text{Tr} \left( \Lambda_{T_k}^2 \left( \frac{1}{\eta_k} \Lambda_{H_k}^H \Lambda_{E_k} \Lambda_{F_k} + I_{p_k} \right)^{-1} \right)
\]

s.t.

\[
\frac{1}{\eta_k} \text{Tr} (A_{H_k}^H A_{F_k}) = P_{T,k}
\]

(118)

where the constant part \(c_k\) is neglected. For any two \(M \times M\) positive semidefinite Hermitian matrices \(A\) and \(B\), we have \(\text{Tr}(AB) \geq \sum_i \lambda_i(A) \lambda_{M-i+1}(B)\), where \(\lambda_i(Z)\) denotes the ith largest eigenvalue of the matrix \(Z\) [38]. Together with the fact that elements of the diagonal matrix \(\Lambda_{T_k}\) are in decreasing order, the objective function of (118) is minimized when \(A_{H_k}^H A_{F_k}/\eta_k + I_{N_k}(b)\) is a diagonal matrix with the diagonal elements in decreasing order. The objective function can be rewritten as

\[
\text{Tr} \left( \Lambda_{T_k}^2 \left( \frac{1}{\eta_k} \Lambda_{H_k}^H \Lambda_{E_k} \Lambda_{F_k} + I_{N_k} \right)^{-1} \right) = d^T \left( \Lambda_{T_k}^2 \right) d \left( \frac{1}{\eta_k} \Lambda_{H_k}^H \Lambda_{E_k} \Lambda_{F_k} + I_{N_k} \right)^{-1}
\]

\[
\overset{\Delta}{=} f(b)
\]

(119)

where \(d(Z)\) denotes the vector which consists of the main diagonal elements of the matrix \(Z\).

It follows that \(f(b)\) is a Schur-concave function of \(b\) [38, 3.H.3]. Then, based on [15, Theorem 1], the optimal \(A_{F_k}\) has the following structure:

\[
A_{F_k,\text{opt}} = \begin{bmatrix}
A_{F_k,\text{opt}}^1 & 0_{N_k \times p_k} & -N_k \\
0_{p_k \times N_k} & 0_{N_k \times p_k} & -N_k
\end{bmatrix}
\]

(120)

where \(A_{F_k,\text{opt}}^1\) 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 \(A_{G_k,\text{opt}}\) in (112), the structure of the optimal \(A_{G_k,\text{opt}}\) is given by

\[
A_{G_k,\text{opt}} = \begin{bmatrix}
A_{G_k,\text{opt}}^1 & 0_{N_k \times q_k} & -N_k \\
0_{p_k \times N_k} & 0_{N_k \times q_k} & -N_k
\end{bmatrix}
\]

(121)

where \(A_{G_k,\text{opt}}^1\) 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. Agustín, “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-optimization 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 relay systems,” IEEE Trans. Signal Process., vol. 56, no. 2, pp. 727–738, Feb. 2008.

[6] A. S. Belghani, R. Merched, and A. J. Eltawil, “Optimizations of a MIMO relay system,” IEEE Wireless Commun., vol. 24, no. 1, pp. 50–56, Feb. 2007.

[7] B. K. Ghoshdeep, 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 MIMO-OFDM systems with optimal training sequences,” IEEE Trans. Signal Process., vol. 58, no. 10, pp. 5062–5073, Oct. 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] H. J. Hammerstrom and A. Witte, “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, B. C. 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 robustTx-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 multiuser 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. Doehler, and A. H. Aghvami, “Effect of channel uncertainty 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 design 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. Mirm 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.

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

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

[37] S. Kay, Fundamentals 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/