The HKU Scholars Hub   The University of Hong Kong   香港大學學術庫

Title	Maximum mutual information design for amplify-and-forward multi-hop MIMO relaying systems under channel uncertainties
Author(s)	Xing, C; Fei, Z; Ma, S; Kuang, J; Wu, YC
Citation	The 2012 IEEE Wireless Communications and Networking Conference (WCNC), Paris, France, 1-4 April 2012. In IEEE Wireless Communications and Networking Conference Proceedings, 2012, p. 781-786
Issued Date	2012
URL	http://hdl.handle.net/10722/165232
Rights	IEEE Wireless Communications and Networking Conference Proceedings. Copyright © IEEE Communications Society.

Maximum Mutual Information Design for Amplify-and-Forward Multi-Hop MIMO Relaying Systems under Channel Uncertainties

Chengwen Xing[†], Zesong Fei[†], Shaodan Ma[§], Jingming Kuang[†], and Yik-Chung Wu[‡]

[†]School of Information and Electronics, Beijing Institute of Technology, Beijing, China

Email: xingchengwen@gmail.com {zesongfei, jmkuang}@bit.edu.cn

[§] Department of Electrical and Computer Engineering, University of Macau, Macau

Email: shaodanma@umac.mo

[‡] Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong

Email: ycwu@eee.hku.hk

Abstract—In this paper, we investigate maximum mutual information design for multi-hop amplify-and-forward (AF) multiple-input multiple-out (MIMO) relaying systems with imperfect channel state information, i.e., Gaussian distributed channel estimation errors. The robust design is formulated as a matrix-variate optimization problem. Exploiting the elegant properties of Majorization theory and matrix-variate functions, the optimal structures of the forwarding matrices at the relays and precoding matrix at the source are derived. Based on the derived structures, a water-filling solution is proposed to solve the remaining unknown variables.

I. Introduction

It is well-established that the deployment of relays can enhance the coverage of base station and improve the quality of wireless links [1]. With these benefits, cooperative communication has been adopted as a standard transmission technique in the future wireless standards such as LTE-A, etc. Specifically, amplify-and-forward (AF) multiple-in multiple-out (MIMO) relaying technology has received a lot of attention of academic and industrial researchers [2]–[4].

Transceiver design for AF MIMO relaying systems, of great importance for system performance, is widely studied [2]–[5]. There are various design criteria with different goals. The most commonly used criteria are capacity maximization [2], [4] and data mean-square-error (MSE) minimization [3], [4]. In most of the previous works, channel state information (CSI) is assumed to be perfectly known. In practice, channel estimation errors are inevitable because of limited length of training sequences. To mitigate the negative effect of channel estimation errors, robust designs are necessary for practical applications.

For linear channel estimators, the estimation errors can be shown to be random with Gaussian distribution [6]. Recently, Bayesian robust transceiver design minimizing weighted MSE for dual-hop AF relaying systems under Gaussian channel

estimation errors has been reported in [6], [7]. On the other hand, information rate maximization for dual-hop AF relaying systems with imperfect channel state information (CSI) has been addressed in [8].

In this paper, we take a further step. Maximization of mutual information for **multi-hop** AF MIMO relaying systems is investigated. Taking the Gaussian distributed channel errors into account and based on Bayesian philosophy, the robust design is formulated into a a matrix-variate optimization problem. Using majorization theory and the elegant properties of matrix-monotone functions, the optimal structures of the precoder at source, multiple forwarding matrices at the relays are derived first. Then a water-filling solution is given for the remaining unknown variables.

The following notations are used throughout this paper. Boldface lowercase letters denote vectors, while boldface uppercase letters denote matrices. The notation $\mathbf{Z}^{\mathrm{H}}$ denotes the Hermitian of the matrix $\mathbf{Z}$, 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. The notation $\mathbf{Z}^{1/2}$ is the Hermitian square root of the positive semidefinite matrix $\mathbf{Z}$, such that $\mathbf{Z}^{1/2}\mathbf{Z}^{1/2} = \mathbf{Z}$ and $\mathbf{Z}^{1/2}$ is also a Hermitian matrix. For two Hermitian matrices, $\mathbf{C} \succeq \mathbf{D}$ means that $\mathbf{C} - \mathbf{D}$ is a positive semi-definite matrix.

II. System Model

In this paper, a multi-hop AF MIMO relaying system is considered. There is one source with N_1 antennas, one destination with M_K antennas and $K - 1$ relays. The kth relay is equipped with M_k receive antennas and N_{k+1} transmit antennas. At the source, a $N \times 1$ data vector s with covariance matrix $\mathbf{R_s} = \mathbb{E}\{\mathbf{ss}^{\mathrm{H}}\} = \mathbf{I}_N$ is transmitted through a precoder matrix $\mathbf{P}_1$. The received signal $\mathbf{x}_1$ at the first relay is $\mathbf{x}_1 = \mathbf{H}_1\mathbf{P}_1\mathbf{s} + \mathbf{n}_1$, where $\mathbf{H}_1$ is the MIMO channel matrix between the source and the first relay, and $\mathbf{n}_1$ is the additive Gaussian noise vector at the first relay with zero mean and covariance matrix $\mathbf{R}_{n_1} = \sigma_{n_1}^2\mathbf{I}_{M_1}$.

At the first relay, the received signal $\mathbf{x}_1$ is first multiplied by a forwarding matrix $\mathbf{P}_2$ and then the resultant signal is

transmitted to the second relay. The received signal $\mathbf{x}_2$ at the second relay is given by

$$\mathbf{x}_2 = \mathbf{H}_2\mathbf{P}_2\mathbf{x}_1 + \mathbf{n}_2 = \mathbf{H}_2\mathbf{P}_2\mathbf{H}_1\mathbf{P}_1\mathbf{s} + \mathbf{H}_2\mathbf{P}_2\mathbf{n}_1 + \mathbf{n}_2, \quad (1)$$

where $\mathbf{H}_2$ is the MIMO channel matrix between the first relay and the second relay, and $\mathbf{n}_2$ is the additive Gaussian noise vector at the second relay with zero mean and covariance matrix $\mathbf{R}_{n_2} = \sigma_{n_2}^2\mathbf{I}_{M_2}$. Similarly, the received signal at k^{th} relay can be written as

$$\mathbf{x}_k = \mathbf{H}_k\mathbf{P}_k\mathbf{x}_{k-1} + \mathbf{n}_k \quad (2)$$

where $\mathbf{H}_k$ is the channel for the kth hop, and $\mathbf{n}_k$ is the additive Gaussian noise with zero mean and covariance matrix $\mathbf{R}_{n_k} = \sigma_{n_k}^2\mathbf{I}_{M_k}$.

Finally, for a K-hop AF MIMO relaying system, the received signal at the destination is

$$\mathbf{y} = \left[\prod_{k=1}^{K}\mathbf{H}_k\mathbf{P}_k\right]\mathbf{s} + \sum_{k=1}^{K-1}\left\{\left[\prod_{l=k+1}^{K}\mathbf{H}_l\mathbf{P}_l\right]\mathbf{n}_k\right\} + \mathbf{n}_K, \quad (3)$$

where $\prod_{k=1}^{K}\mathbf{Z}_k$ denotes $\mathbf{Z}_K \times \cdots \times \mathbf{Z}_1$. In order to guarantee the transmitted data $\mathbf{s}$ can be recovered at the destination, it is assumed that N_k and M_k are greater than or equal to N [7].

In practical systems, because of limited length of training sequences, channel estimation errors are inevitable [12]. With channel estimation errors, we can write

$$\mathbf{H}_k = \bar{\mathbf{H}}_k + \Delta\mathbf{H}_k, \quad (4)$$

where $\bar{\mathbf{H}}_k$ is the estimated channel in the kth hop and $\Delta\mathbf{H}_k$ is the corresponding channel estimation error whose elements are zero mean Gaussian random variables. Moreover, the $M_k \times N_k$ matrix $\Delta\mathbf{H}_k$ can be decomposed using the widely used Kronecker model $\Delta\mathbf{H}_k = \Sigma_k^{1/2}\mathbf{H}_{W,k}\Psi_k^{1/2}$ [6], [7], [9], [10]. The elements of the $M_k \times N_k$ matrix $\mathbf{H}_{W,k}$ are independent and identically distributed (i.i.d.) Gaussian random variables with zero mean and unit variance. The specific formulas of the row correlation matrix Σ_k and the column correlation matrix Ψ_k are determined by the training sequences and channel estimators being used [6].

At the destination, a linear equalizer $\mathbf{G}$ is employed to detect the desired data vector $\mathbf{s}$. The resulting data MSE matrix equals to $\Phi(\mathbf{G}) = \mathbb{E}\{(\mathbf{G}\mathbf{y} - \mathbf{s})(\mathbf{G}\mathbf{y} - \mathbf{s})^{\text{H}}\}$, where the expectation is taken with respect to random data, channel estimation errors, and noise. Following a similar derivation in dual-hop systems [7], the MSE matrix is derived to be

$$\Phi(\mathbf{G})$$
$$= \mathbb{E}\{(\mathbf{G}\mathbf{y} - \mathbf{s})(\mathbf{G}\mathbf{y} - \mathbf{s})^{\text{H}}\}$$
$$= \mathbf{G}[\bar{\mathbf{H}}_K\mathbf{P}_K\mathbf{R}_{\mathbf{x}_{K-1}}\mathbf{P}_K^{\text{H}}\bar{\mathbf{H}}_K^{\text{H}} + \text{Tr}(\mathbf{P}_K\mathbf{R}_{\mathbf{x}_{K-1}}\mathbf{P}_K^{\text{H}}\Psi_K)\Sigma_K$$
$$+ \mathbf{R}_{n_K}]\mathbf{G}^{\text{H}} + \mathbf{I}_N - \left[\prod_{k=1}^{K}\bar{\mathbf{H}}_k\mathbf{P}_k\right]^{\text{H}}\mathbf{G}^{\text{H}} - \mathbf{G}\left[\prod_{k=1}^{K}\bar{\mathbf{H}}_k\mathbf{P}_k\right], \quad (5)$$

where the received signal covariance matrix $\mathbf{R}_{\mathbf{x}_k}$ at the kth relay satisfies the following recursive formula

$$\mathbf{R}_{\mathbf{x}_k} = \bar{\mathbf{H}}_k\mathbf{P}_k\mathbf{R}_{\mathbf{x}_{k-1}}\mathbf{P}_k^{\text{H}}\bar{\mathbf{H}}_k^{\text{H}} + \text{Tr}(\mathbf{P}_k\mathbf{R}_{\mathbf{x}_{k-1}}\mathbf{P}_k^{\text{H}}\Psi_k)\Sigma_k + \mathbf{R}_{n_k}, \quad (6)$$

and $\mathbf{R}_{\mathbf{x_0}} = \mathbf{R}_\mathbf{s} = \mathbf{I}_N$ represents the signal covariance matrix at the source.

III. PROBLEM FORMULATION

Maximizing mutual information is one of the most important and widely used performance metric for transceiver design. Denoting the received pilot for the channel estimation as $\mathbf{r}$, the channel capacity between the source and destination is $I(\mathbf{s};\mathbf{y}|\mathbf{r})$ [13]. Unfortunately to the best of our knowledge, the exact capacity of MIMO channels with channel estimation errors is still open even for point-to-point MIMO systems [10], [13]. To proceed, a lower bound of capacity is usually exploited [8]

$$-\log|\Phi(\mathbf{G})| \leq I(\mathbf{s};\mathbf{y}|\mathbf{r}), \quad (7)$$

where the equality holds when the CSI is perfectly known [2], [11]. For imperfect CSI, the tightness of this bound is extensively investigated in [10], [13]. Based on this lower bound, the robust transceiver design maximizing mutual information can be replaced by minimizing the following objective function [10]

$$\min_{\mathbf{P}_k,\mathbf{G}} \quad \log|\Phi(\mathbf{G})|$$
$$\text{s.t.} \quad \text{Tr}(\mathbf{P}_k\mathbf{R}_{\mathbf{x}_{k-1}}\mathbf{P}_k^{\text{H}}) \leq P_k, \quad k = 1, \cdots, K \quad (8)$$

where the objective function $\log|\Phi(\mathbf{G})|$ is a real-valued matrix-variate function with $\Phi(\mathbf{G})$ as its argument. Furthermore, $\log|\Phi(\mathbf{G})|$ is a matrix-monotone increasing function [14].

For (8), there is no constraint on the equalizer $\mathbf{G}$. We can differentiate the trace of (5) with respect to $\mathbf{G}$ and obtain the LMMSE equalizer

$$\mathbf{G}_{\text{LMMSE}} = \left[\prod_{k=1}^{K}\bar{\mathbf{H}}_k\mathbf{P}_k\right]^{\text{H}}[\bar{\mathbf{H}}_K\mathbf{P}_K\mathbf{R}_{\mathbf{x}_{K-1}}\mathbf{P}_K^{\text{H}}\bar{\mathbf{H}}_K^{\text{H}}$$
$$+ \text{Tr}(\mathbf{P}_K\mathbf{R}_{\mathbf{x}_{K-1}}\mathbf{P}_K^{\text{H}}\Psi_K)\Sigma_K + \mathbf{R}_{n_K}]^{-1}, \quad (9)$$

with the following property

$$\Phi(\mathbf{G}_{\text{LMMSE}}) \preceq \Phi(\mathbf{G}). \quad (10)$$

Because $\log|\Phi(\mathbf{G})|$ is a matrix-monotone increasing function, (10) implies that $\mathbf{G}_{\text{LMMSE}}$ minimizes the objective function in (8). Substituting the optimal equalizer of (9) into $\Phi(\mathbf{G})$ in (5), $\Phi(\mathbf{G})$ equals to

$$\Phi_{\text{MSE}} = \mathbf{I}_N - \left[\prod_{k=1}^{K}\bar{\mathbf{H}}_k\mathbf{P}_k\right]^{\text{H}}[\bar{\mathbf{H}}_K\mathbf{P}_K\mathbf{R}_{\mathbf{x}_{K-1}}\mathbf{P}_K^{\text{H}}\bar{\mathbf{H}}_K^{\text{H}}$$
$$+ \text{Tr}(\mathbf{P}_K\mathbf{R}_{\mathbf{x}_{K-1}}\mathbf{P}_K^{\text{H}}\Psi_K)\Sigma_K + \mathbf{R}_{n_K}]^{-1}\left[\prod_{k=1}^{K}\bar{\mathbf{H}}_k\mathbf{P}_k\right]. \quad (11)$$

For multi-hop AF MIMO relaying systems, the received signal at k^{th} relay depends on the forwarding matrices at all preceding relays, making the power allocations at different relays couples with each other (as seen in the constraints of (8)), and thus the problem (8) difficult to solve. In order to simplify the problem, we define the following new variable in terms of $\mathbf{P}_k$:

$$\mathbf{F}_k \triangleq \mathbf{P}_k \mathbf{K}_{\mathbf{F}_{k-1}}^{1/2}$$
$$\times \underbrace{(\mathbf{K}_{\mathbf{F}_{k-1}}^{-1/2} \bar{\mathbf{H}}_{k-1} \mathbf{F}_{k-1} \mathbf{F}_{k-1}^{\mathrm{H}} \bar{\mathbf{H}}_{k-1}^{\mathrm{H}} \mathbf{K}_{\mathbf{F}_{k-1}}^{-1/2} + \mathbf{I}_{M_{k-1}})^{1/2}}_{\triangleq \mathbf{\Pi}_{k-1}},$$
(12)

where $\mathbf{K}_{\mathbf{F}_k} \triangleq \mathrm{Tr}(\mathbf{F}_k \mathbf{F}_k^{\mathrm{H}} \mathbf{\Psi}_k) \mathbf{\Sigma}_k + \sigma_{n_k}^2 \mathbf{I}_{M_k}$. Notice that $\mathbf{F}_1 = \mathbf{P}_1$. With the new variable, the MSE matrix $\mathbf{\Phi}_{\text{MSE}}$ is reformulated as

$$\mathbf{\Phi}_{\text{MSE}} = \mathbf{I}_N - \left[\prod_{k=1}^{K} \mathbf{\Pi}_k^{-1/2} \mathbf{K}_{\mathbf{F}_k}^{-1/2} \bar{\mathbf{H}}_k \mathbf{F}_k \right]^{\mathrm{H}}$$
$$\times \left[\prod_{k=1}^{K} \underbrace{\mathbf{\Pi}_k^{-1/2} \mathbf{K}_{\mathbf{F}_k}^{-1/2} \bar{\mathbf{H}}_k \mathbf{F}_k}_{\triangleq \mathbf{A}_k} \right]$$
$$= \mathbf{I}_N - \mathbf{A}_1^{\mathrm{H}} \cdots \mathbf{A}_K^{\mathrm{H}} \mathbf{A}_K \cdots \mathbf{A}_1. \qquad (13)$$

Meanwhile, with the new variables $\mathbf{F}_k$, the corresponding power constraint in the kth hop can now be rewritten as

$$\mathrm{Tr}(\mathbf{F}_k \mathbf{F}_k^{\mathrm{H}}) \le P_k. \qquad (14)$$

It is obvious that with the new variables $\mathbf{F}_k$, the constraints become independent of each other. Putting (13) and (14) into (8), the transceiver design problem can be reformulated as

$$\textbf{P 1:} \quad \min_{\mathbf{F}_k, \mathbf{Q}_k} \quad \log|\mathbf{I}_N - \mathbf{\Theta}|$$
$$\text{s.t.} \quad \mathbf{\Theta} = \mathbf{A}_1^{\mathrm{H}} \cdots \mathbf{A}_K^{\mathrm{H}} \mathbf{A}_K \cdots \mathbf{A}_1$$
$$\mathrm{Tr}(\mathbf{F}_k \mathbf{F}_k^{\mathrm{H}}) \le P_k, \quad k = 1, \cdots, K. \qquad (15)$$

From the definition of $\mathbf{A}_k$ in (13) and noticing that $\mathbf{K}_{\mathbf{F}_k} = \mathrm{Tr}(\mathbf{F}_k \mathbf{F}_k^{\mathrm{H}} \mathbf{\Psi}_k) \mathbf{\Sigma}_k + \sigma_{n_k}^2 \mathbf{I}_{M_k}$, it can be seen that $\mathbf{F}_k$ appears at multiple positions in the objective function. Therefore, the optimization problem is much more complicated than the counterpart with prefect CSI. Indeed, as demonstrated by existing works, robust transceiver design for point-to-point or dual-hop relaying MIMO systems is much more complicated and challenging than its counterpart with perfect CSI [6], [9], [10].

IV. OPTIMAL STRUCTURE OF ROBUST TRANSCEIVER

The objective function of (15) can be directly replaced by a function of $\boldsymbol{\lambda}(\mathbf{\Theta}) = [\lambda_1(\mathbf{\Theta}), \cdots, \lambda_N(\mathbf{\Theta})]^{\mathrm{T}}$, where the symbol $\lambda_i(\mathbf{Z})$ represents the ith largest eigenvalue of $\mathbf{Z}$. Thus the optimization problem becomes

$$\textbf{P 2:} \quad \min_{\mathbf{F}_k} \quad \boldsymbol{g}[\boldsymbol{\lambda}(\mathbf{\Theta})] = \sum_{i=1}^{N} \log(1 - \lambda_i(\mathbf{\Theta}))$$
$$\text{s.t.} \quad \mathbf{\Theta} = \mathbf{A}_1^{\mathrm{H}} \cdots \mathbf{A}_K^{\mathrm{H}} \mathbf{A}_K \cdots \mathbf{A}_1$$
$$\mathrm{Tr}(\mathbf{F}_k \mathbf{F}_k^{\mathrm{H}}) \le P_k. \qquad (16)$$

where $\mathbf{A}_k$'s are defined in (13). In order to further simplify the optimization problem, we make use of the following two additional properties.

Property 1: As $\boldsymbol{g}(\bullet)$ is a decreasing and Schur-concave function, the objective function in **P 2** satisfies

$$\boldsymbol{g}(\boldsymbol{\lambda}(\mathbf{\Theta})) \ge \boldsymbol{g}(\underbrace{[\gamma_1(\mathbf{\Theta}) \ \cdots \gamma_N(\mathbf{\Theta})]^{\mathrm{T}}}_{\triangleq \boldsymbol{\gamma}(\mathbf{\Theta})}) \qquad (17)$$

with $\gamma_i(\mathbf{\Theta}) \triangleq \lambda_i(\mathbf{A}_K^{\mathrm{H}} \mathbf{A}_K) \lambda_i(\mathbf{A}_{K-1}^{\mathrm{H}} \mathbf{A}_{K-1}) \cdots \lambda_i(\mathbf{A}_1^{\mathrm{H}} \mathbf{A}_1),$ (18)

where the equality in (17) holds when the neighboring $\mathbf{A}_k$'s satisfy

$$\mathbf{V}_{\mathbf{A}_k} = \mathbf{U}_{\mathbf{A}_{k-1}}, \quad k = 2, \cdots, K. \qquad (19)$$

In (19), unitary matrices $\mathbf{U}_{\mathbf{A}_k}$ and $\mathbf{V}_{\mathbf{A}_k}$ are defined based on the singular value decomposition (SVD) $\mathbf{A}_k = \mathbf{U}_{\mathbf{A}_k} \mathbf{\Lambda}_{\mathbf{A}_k} \mathbf{V}_{\mathbf{A}_k}^{\mathrm{H}}$ with $\mathbf{\Lambda}_{\mathbf{A}_k} \searrow$.
Proof: See Appendix A. ∎

Notice that the achievement of equality in **Property 1** does not affect the constraints in **P 1**, as (19) in **Property 1** only needs to introduce several unitary matrices from $\mathbf{F}_k$'s while the constraints of $\mathbf{F}_k$'s in **P 1** are unitary-invariant.

As $\boldsymbol{g}(\bullet)$ is a monotonically decreasing function with respect to its vector argument, the optimal solutions of the optimization problem **P 2** always occur on the boundary $\mathrm{Tr}(\mathbf{F}_k \mathbf{F}_k^{\mathrm{H}}) = P_k$ [7]. Notice that this coincides with intuition, but intuition cannot be used as theoretical basis. Everything should be carefully proved.
Property 2: Defining

$$\eta_{f_k} \triangleq \mathrm{Tr}(\mathbf{F}_k \mathbf{F}_k^{\mathrm{H}} \mathbf{\Psi}_k) \alpha_k + \sigma_{n_k}^2 \qquad (20)$$

with $\alpha_k = \mathrm{Tr}(\mathbf{\Sigma}_k)/M_k$ which is a constant, $\mathrm{Tr}(\mathbf{F}_k \mathbf{F}_k^{\mathrm{H}}) = P_k$ is equivalent to

$$\mathrm{Tr}[\mathbf{F}_k \mathbf{F}_k^{\mathrm{H}} (\alpha_k P_k \mathbf{\Psi}_k + \sigma_{n_k}^2 \mathbf{I}_{N_k})]/\eta_{f_k} = P_k. \qquad (21)$$

Proof: See Appendix B. ∎

Based on **Properties 1** and **2**, the optimal solution of the optimization problem (16) is exactly the optimal solution of the following new optimization problem with different constraints

$$\textbf{P 3:} \quad \min_{\mathbf{F}_k} \quad \boldsymbol{g}[\boldsymbol{\gamma}(\mathbf{\Theta})]$$
$$\text{s.t.} \quad \mathrm{Tr}[\mathbf{F}_k \mathbf{F}_k^{\mathrm{H}} (\alpha_k P_k \mathbf{\Psi}_k + \sigma_{n_k}^2 \mathbf{I}_{N_k})]/\eta_{f_k} = P_k$$
$$\mathbf{\Theta} = \mathbf{A}_1^{\mathrm{H}} \cdots \mathbf{A}_K^{\mathrm{H}} \mathbf{A}_K \cdots \mathbf{A}_1$$
$$\mathbf{V}_{\mathbf{A}_k} = \mathbf{U}_{\mathbf{A}_{k-1}}, \quad k = 2, \cdots, K. \qquad (22)$$

Noticing that $\boldsymbol{g}(\bullet)$ is a monotonically decreasing function, solving **P 3** gives the following structure for the optimal solution.

Conclusion 1: Defining unitary matrices $\mathbf{U}_{\mathcal{H}_k}$ and $\mathbf{V}_{\mathcal{H}_k}$ based on the following SVD

$$(\mathbf{K}_{\mathbf{F}_k}/\eta_{f_k})^{-1/2}\bar{\mathbf{H}}_k(\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2 \mathbf{I}_{N_k})^{-1/2} = \mathbf{U}_{\mathcal{H}_k}\boldsymbol{\Lambda}_{\mathcal{H}_k}\mathbf{V}_{\mathcal{H}_k}^{\mathrm{H}}$$

with $\boldsymbol{\Lambda}_{\mathcal{H}_k}\searrow$ and $\mathbf{U}_{\mathcal{H}_0} = \mathbf{U}_{\mathrm{Arb}}$, $\qquad(23)$

when $\boldsymbol{\Psi}_k \propto \mathbf{I}$ or $\boldsymbol{\Sigma}_k \propto \mathbf{I}$, the optimal solutions of the optimization problem (22) have the following structure

$$\mathbf{F}_{k,\mathrm{opt}} = \sqrt{\boldsymbol{\xi}_k(\boldsymbol{\Lambda}_{\mathcal{F}_k})}(\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2 \mathbf{I}_{N_k})^{-1/2}\mathbf{V}_{\mathcal{H}_k,N}\boldsymbol{\Lambda}_{\mathcal{F}_k}$$
$$\times \mathbf{U}_{\mathcal{H}_{k-1},N}^{\mathrm{H}}, \qquad(24)$$

where $\mathbf{V}_{\mathcal{H}_k,N}$ and $\mathbf{U}_{\mathcal{H}_k,N}$ are the matrices consisting of the first N columns of $\mathbf{V}_{\mathcal{H}_k}$ and $\mathbf{U}_{\mathcal{H}_k}$, respectively, $\mathbf{U}_{\mathrm{Abr}}$ is an arbitrary $N \times N$ unitary matrix and $\boldsymbol{\Lambda}_{\mathcal{F}_k}$ is a $N \times N$ unknown diagonal matrix. The scalar $\boldsymbol{\xi}_k(\boldsymbol{\Lambda}_{\mathcal{F}_k})$ is a function of $\boldsymbol{\Lambda}_{\mathcal{F}_k}$ and equals to

$$\boldsymbol{\xi}_k(\boldsymbol{\Lambda}_{\mathcal{F}_k}) = \sigma_{n_k}^2/\{1 - \alpha_k \mathrm{Tr}[\mathbf{V}_{\mathcal{H}_k,N}^{\mathrm{H}}(\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2 \mathbf{I}_{N_k})^{-1/2}$$
$$\times \boldsymbol{\Psi}_k(\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2 \mathbf{I}_{N_k})^{-1/2}\mathbf{V}_{\mathcal{H}_k,N}\boldsymbol{\Lambda}_{\mathcal{F}_k}^2]\} = \eta_{f_k}.$$
$$(25)$$

Proof: See Appendix C. ∎

In the optimal structure given by (24), the scalar variable $\boldsymbol{\xi}_k(\boldsymbol{\Lambda}_{\mathcal{F}_k})$ is only a function of the matrix $\boldsymbol{\Lambda}_{\mathcal{F}_k}$ and therefore the only unknown variable in (24) is $\boldsymbol{\Lambda}_{\mathcal{F}_k}$. The remaining unknown diagonal elements of $\boldsymbol{\Lambda}_{\mathcal{F}_k}$ can be obtained by water-filling alike solution as discussed in the next section.

Remark:

(a) From **_Conclusion 1_**, it can be seen that with channel estimation errors, the optimal transceiver structure is totally different from its counterpart with perfect CSI [5]. The results given in [5] cannot be transferred to our proposed solution. Even for point-to-point MIMO systems, the robust transceiver design with column correlations is derived lately based on a complicated discussion from KKT conditions [10].

(b) We also want to highlight that the design with perfect CSI is a special case of the proposed robust design. With perfect CSI, the main difference between our work and the pioneering works [4], [5] is that we do not need a common rank constraint on the all transceiver matrices, which is a prior condition of the proofs in [4], [5].

(c) The formulation of the optimization problem **P 1** reveals the relationship between AF MIMO relaying systems and traditional point-to-point MIMO systems. It is obvious that a point-to-point MIMO system is a special case of the considered AF MIMO relaying system.

(d) The formulations given by this paper can be directly extended to multi-user cases by changing the transceiver matrices into block diagonal matrices.

V. COMPUTATIONS OF $\boldsymbol{\Lambda}_{\mathcal{F}_k}$

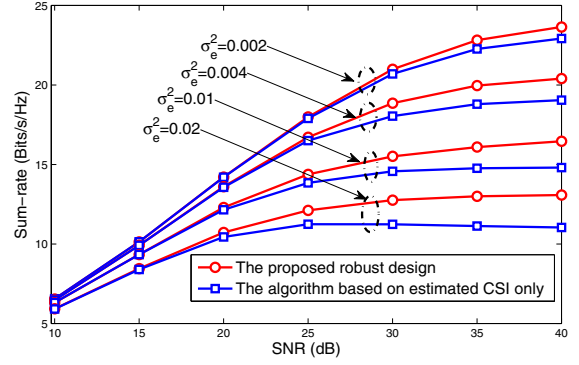Based on the optimal structure given by **_Conclusion 1_**, the optimization problem (22) can be written as

Fig. 1. Sum-rates of a three-hop AF relaying system when $\alpha = 0.6$ and $\beta = 0$.

$$\min_{f_{k,i}} \quad \sum_{i=1}^{N}\log\left(1 - \frac{\prod_{k=1}^{K}f_{k,i}^2 h_{k,i}^2}{\prod_{k=1}^{K}(f_{k,i}^2 h_{k,i}^2 + 1)}\right)$$
$$\text{s.t.} \quad \sum_{i=1}^{N}f_{k,i}^2 = P_k. \qquad(26)$$

Because this optimization problem is nonconvex, there is no closed-form solution [15] and therefore an iterative water-filling algorithm can be used to solve for $f_{k,i}$ with convergence guaranteed. More specifically, when $f_{l,i}$'s are fixed with $l \neq k$, $f_{k,i}$ is computed as

$$f_{k,i}^2 = \frac{1}{h_{k,i}^2}\left(\frac{-a_{k,i} + \sqrt{a_{k,i}^2 + 4(1-a_{k,i})a_{k,i}h_{k,i}^2/\mu_k}}{2(1-a_{k,i})} - 1\right)^+$$

with $a_{k,i} = \prod_{l\neq k} f_{l,i}^2 h_{l,i}^2/(f_{l,i}^2 h_{l,i}^2 + 1)$ $\qquad(27)$

where μ_k is the Lagrange multiplier which makes $\sum_{i=1}^{N}f_{k,i}^2 = P_k$ hold.

VI. SIMULATION RESULTS AND DISCUSSIONS

For the purpose of comparison, the algorithms based on the estimated channel only (without taking the channel estimation errors into account) are also simulated. In the following, we consider a three-hop AF MIMO relaying system where all nodes are equipped with 4 antennas. Furthermore, the estimation error correlation matrices are chosen as the popular exponential model [7] i.e., $[\boldsymbol{\Psi}_k]_{i,j} = \sigma_e^2 \alpha^{|i-j|}$ and $[\boldsymbol{\Sigma}_k]_{i,j} = \beta^{|i-j|}$. are chosen based on the popular exponential model [7], [9], and σ_e^2 denotes the estimation error variance. The estimated channels $\bar{\mathbf{H}}_k$'s, are generated based on the following complex Gaussian distributions $\bar{\mathbf{H}}_k \sim \mathcal{CN}_{M_k,N_k}(\mathbf{0}_{M_k,N_k},(1-\sigma_e^2)/\sigma_e^2\boldsymbol{\Sigma}_k \otimes \boldsymbol{\Psi}_k^{\mathrm{T}})$, such that channel realizations $\mathbf{H}_k = \bar{\mathbf{H}}_k + \Delta\mathbf{H}_k$ have unit variance. We define the signal-to-noise ratio (SNR) for the kth link as $P_k/\sigma_{n_k}^2$.

Fig. 1 shows the sum-rates resulting from the proposed algorithm and the algorithm based on estimated CSI only with different SNRs (SNR = $P_k/\sigma_{n_k}^2$). It is assumed that the SNRs at various hops are the same. The correlation coefficients in

the channel estimation errors are taken as $\alpha = 0.6$ and $\beta = 0$. It can be seen that the performance of the proposed algorithm is better than that of the corresponding algorithm based on estimated CSI only. Furthermore, as the estimation errors increase, the performance gap between the two algorithms becomes larger.

VII. CONCLUSIONS

Maximum information design for multi-hop AF MIMO relaying systems under Gaussian distributed channel estimation errors was considered. Using majorization theory and the properties of matrix-monotone functions, the optimal structure of transceivers was derived. Then, the transceiver design problem was significantly simplified and the remaining unknowns were obtained by an iterative water-filling solution. The performance advantage of the proposed design was assessed by numerical results.

APPENDIX A
PROOF OF PROPERTY 1

First notice that for two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ with compatible dimension, $\lambda_i(\boldsymbol{AB}) = \lambda_i(\boldsymbol{BA})$ [14, 9.A.1.a]. Together with the fact that for two positive semi-definite matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, $\prod_{i=1}^{k} \lambda_i(\boldsymbol{AB}) \leq \prod_{i=1}^{k} \lambda_i(\boldsymbol{A})\lambda_i(\boldsymbol{B})$ [14, 9.H.1.a], we have

$$\prod_{i=1}^{k} \lambda_i(\boldsymbol{A}_1^{\mathrm{H}} \cdots \boldsymbol{A}_K^{\mathrm{H}} \boldsymbol{A}_K \cdots \boldsymbol{A}_1)$$

$$\leq \prod_{i=1}^{k} \lambda_i(\boldsymbol{A}_2^{\mathrm{H}} \cdots \boldsymbol{A}_K^{\mathrm{H}} \boldsymbol{A}_K \cdots \boldsymbol{A}_2)\lambda_i(\boldsymbol{A}_1 \boldsymbol{A}_1^{\mathrm{H}}). \quad (28)$$

Repeating this process we have the following inequality

$$\prod_{i=1}^{k} \lambda_i(\boldsymbol{A}_1^{\mathrm{H}} \cdots \boldsymbol{A}_K^{\mathrm{H}} \boldsymbol{A}_K \cdots \boldsymbol{A}_1)$$

$$\leq \prod_{i=1}^{k} \underbrace{\lambda_i(\boldsymbol{A}_K^{\mathrm{H}} \boldsymbol{A}_K)\lambda_i(\boldsymbol{A}_{K-1}^{\mathrm{H}} \boldsymbol{A}_{K-1}) \cdots \lambda_i(\boldsymbol{A}_1^{\mathrm{H}} \boldsymbol{A}_1)}_{\triangleq \gamma_i(\boldsymbol{\Theta})}. \quad (29)$$

This is an important and useful conclusion for transceiver design for AF MIMO relaying systems.

Based on (29) and **5.A.2.b** in [14], we directly have

$$\boldsymbol{\lambda}(\boldsymbol{A}_1^{\mathrm{H}} \cdots \boldsymbol{A}_K^{\mathrm{H}} \boldsymbol{A}_K \cdots \boldsymbol{A}_1) \prec_w [\gamma_1(\boldsymbol{\Theta}) \ \gamma_2(\boldsymbol{\Theta}) \ \cdots \gamma_N(\boldsymbol{\Theta})]^{\mathrm{T}} \quad (30)$$

with equality holds if and only if the neighboring $\boldsymbol{A}_k$'s satisfy

$$\mathbf{V}_{\boldsymbol{A}_k} = \mathbf{U}_{\boldsymbol{A}_{k-1}}, \quad k = 2, \cdots, K \quad (31)$$

where $\mathbf{U}_{\boldsymbol{A}_k}$ and $\mathbf{V}_{\boldsymbol{A}_k}$ are defined based on the SVD $\boldsymbol{A}_k = \mathbf{U}_{\boldsymbol{A}_k} \boldsymbol{\Lambda}_{\boldsymbol{A}_k} \mathbf{V}_{\boldsymbol{A}_k}^{\mathrm{H}}$ with $\boldsymbol{\Lambda}_{\boldsymbol{A}_k} \searrow$.

As $\boldsymbol{g}(\bullet)$ is a decreasing and Schur-concave function, it can be concluded that [14]

$$\boldsymbol{g}[\boldsymbol{\lambda}(\boldsymbol{\Theta})] \geq \boldsymbol{g}[\boldsymbol{\gamma}(\boldsymbol{\Theta})], \quad (32)$$

where $\boldsymbol{\gamma}(\boldsymbol{\Theta}) = [\gamma_1(\boldsymbol{\Theta}) \ \cdots \gamma_N(\boldsymbol{\Theta})]^{\mathrm{T}}$, and with equality holds if and only if (31) holds.

APPENDIX B
PROOF OF PROPERTY 2

Defining $\eta_{f_k} = \alpha_k \mathrm{Tr}(\mathbf{F}_k \mathbf{F}_k^{\mathrm{H}} \boldsymbol{\Psi}_k) + \sigma_{n_k}^2$, when $\mathrm{Tr}(\mathbf{F}_k \mathbf{F}_k^{\mathrm{H}}) = P_k$, we can write

$$\begin{aligned}
\eta_{f_k} &= \alpha_k \mathrm{Tr}(\mathbf{F}_k \mathbf{F}_k^{\mathrm{H}} \boldsymbol{\Psi}_k) + \sigma_{n_k}^2 \\
&= \alpha_k \mathrm{Tr}(\mathbf{F}_k \mathbf{F}_k^{\mathrm{H}} \boldsymbol{\Psi}_k) + \sigma_{n_k}^2 \underbrace{\mathrm{Tr}(\mathbf{F}_k \mathbf{F}_k^{\mathrm{H}})/P_k}_{=1} \\
&= \mathrm{Tr}(\mathbf{F}_k \mathbf{F}_k^{\mathrm{H}}(\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2 \mathbf{I}))/P_k, \quad (33)
\end{aligned}$$

based on which it can be directly concluded that

$$\mathrm{Tr}[\mathbf{F}_k \mathbf{F}_k^{\mathrm{H}}(\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2 \mathbf{I})]/\eta_{f_k} = P_k. \quad (34)$$

On the other hand, when $\mathrm{Tr}[\mathbf{F}_k \mathbf{F}_k^{\mathrm{H}}(\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2 \mathbf{I})]/\eta_{f_k} = P_k$ with the definition of η_{f_k} we have

$$\mathrm{Tr}[\mathbf{F}_k \mathbf{F}_k^{\mathrm{H}}(\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2 \mathbf{I})] = \alpha_k P_k \mathrm{Tr}(\mathbf{F}_k \mathbf{F}_k^{\mathrm{H}} \boldsymbol{\Psi}_k) + \sigma_{n_k}^2 P_k, \quad (35)$$

which means $\mathrm{Tr}(\mathbf{F}_k \mathbf{F}_k^{\mathrm{H}}) = P_k$.

APPENDIX C
PROOF OF CONCLUSION 1

Problem reformulation: Note that

$$\gamma_i(\boldsymbol{\Theta}) = \prod_{k=1}^{K} \frac{\lambda_i(\mathbf{F}_k^{\mathrm{H}} \bar{\mathbf{H}}_k^{\mathrm{H}} \mathbf{K}_{\mathbf{F}_k}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_k)}{1 + \lambda_i(\mathbf{F}_k^{\mathrm{H}} \bar{\mathbf{H}}_k^{\mathrm{H}} \mathbf{K}_{\mathbf{F}_k}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_k)}, \quad (36)$$

based on which $\gamma_i(\boldsymbol{\Theta})$ is a function of $\boldsymbol{\lambda}(\mathbf{F}_k^{\mathrm{H}} \bar{\mathbf{H}}_k^{\mathrm{H}} \mathbf{K}_{\mathbf{F}_k}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_k)$. Unfortunately, $\mathbf{F}_k$ appears in multiple positions. In particular, $\mathbf{K}_{\mathbf{F}_k}$ is a function of $\mathbf{F}_k$ which complicates the derivation of optimal solutions. In order to simplify the problem, $\boldsymbol{\lambda}(\mathbf{F}_k^{\mathrm{H}} \bar{\mathbf{H}}_k^{\mathrm{H}} \mathbf{K}_{\mathbf{F}_k}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_k)$ is reformulated as

$$\begin{aligned}
&\boldsymbol{\lambda}(\mathbf{F}_k^{\mathrm{H}} \bar{\mathbf{H}}_k^{\mathrm{H}} \mathbf{K}_{\mathbf{F}_k}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_k) \\
&= \boldsymbol{\lambda}[\tilde{\mathbf{F}}_k^{\mathrm{H}}(\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2 \mathbf{I}_{N_k})^{-1/2} \bar{\mathbf{H}}_k^{\mathrm{H}}(\mathbf{K}_{\mathbf{F}_k}/\eta_{f_k})^{-1/2} \\
&\quad \times \underbrace{(\mathbf{K}_{\mathbf{F}_k}/\eta_{f_k})^{-1/2} \bar{\mathbf{H}}_k(\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2 \mathbf{I}_{N_k})^{-1/2}}_{\triangleq \boldsymbol{\mathcal{H}}_k} \tilde{\mathbf{F}}_k], \quad (37)
\end{aligned}$$

where $\tilde{\mathbf{F}}_k$ is defined as

$$\tilde{\mathbf{F}}_k = 1/\sqrt{\eta_{f_k}}(\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2 \mathbf{I}_{N_k})^{1/2} \mathbf{F}_k. \quad (38)$$

The right hand side of (37) is easier to handle than the left hand side. This is because when $\boldsymbol{\Psi}_k \propto \mathbf{I}_{N_k}$ or $\boldsymbol{\Sigma}_k \propto \mathbf{I}_{M_k}$, $\boldsymbol{\mathcal{H}}_k$ is independent of $\tilde{\mathbf{F}}_k$. In the following, we will prove this in detail.

It is obvious that $\boldsymbol{\mathcal{H}}_k$ being independent of $\tilde{\mathbf{F}}_k$ is equivalent to $\mathbf{K}_{\mathbf{F}_k}/\eta_{f_k}$ being independent of $\tilde{\mathbf{F}}_k$. First consider $\boldsymbol{\Psi}_k \propto \mathbf{I}_{N_k}$, i.e., $\boldsymbol{\Psi}_k = \beta_k \mathbf{I}_{N_k}$. With the definitions of $\mathbf{K}_{\mathbf{F}_k}$ in (12) and η_{f_k} in (20), $\mathbf{K}_{\mathbf{F}_k}/\eta_{f_k}$ equals to

$$\begin{aligned}
&\mathbf{K}_{\mathbf{F}_k}/\eta_{f_k} \\
&= [\beta_k \mathrm{Tr}(\mathbf{F}_k \mathbf{F}_k^{\mathrm{H}}) \boldsymbol{\Sigma}_k + \sigma_{n_k}^2 \mathbf{I}_{M_k}]/[\beta_k \alpha_k \mathrm{Tr}(\mathbf{F}_k \mathbf{F}_k^{\mathrm{H}}) + \sigma_{n_k}^2] \\
&= \frac{\beta_k P_k}{\alpha_k \beta_k P_k + \sigma_{n_k}^2} \boldsymbol{\Sigma}_k + \frac{\sigma_{n_k}^2}{\alpha_k \beta_k P_k + \sigma_{n_k}^2} \mathbf{I}_{M_k}, \quad (39)
\end{aligned}$$

where the second equality is based on the fact that $\mathrm{Tr}(\mathbf{F}_k\mathbf{F}_k^{\mathrm{H}}) = P_k$. On the other hand, when $\boldsymbol{\Sigma}_k \propto \mathbf{I}_{M_k}$ (i.e, $\boldsymbol{\Sigma}_k = \alpha_k\mathbf{I}_{M_k}$), $\mathbf{K}_{\mathbf{F}_k}/\eta_{f_k}$ equals to

$$
\begin{aligned}
&\mathbf{K}_{\mathbf{F}_k}/\eta_{f_k} \\
&= [\alpha_k\mathrm{Tr}(\mathbf{F}_k\mathbf{F}_k^{\mathrm{H}}\boldsymbol{\Psi}_k)\mathbf{I}_{N_k} + \sigma_{n_k}^2\mathbf{I}_{N_k}]/[\alpha_k[\mathrm{Tr}(\mathbf{F}_k\mathbf{F}_k^{\mathrm{H}}\boldsymbol{\Psi}_k) + \sigma_{n_k}^2] \\
&= \mathbf{I}_{N_k}.
\end{aligned} \tag{40}
$$

Therefore, when $\boldsymbol{\Psi}_k \propto \mathbf{I}_{N_k}$ or $\boldsymbol{\Sigma}_k \propto \mathbf{I}_{M_k}$, $\mathbf{K}_{\mathbf{F}_k}/\eta_{f_k}$ is independent of $\tilde{\mathbf{F}}_k$.

Using the substitution (38), the optimization problem (22) is reformulated as

$$
\begin{aligned}
\min_{\tilde{\mathbf{F}}_k} \quad & g[\boldsymbol{\gamma}(\boldsymbol{\Theta})] \\
\mathrm{s.t.} \quad & \gamma_i(\boldsymbol{\Theta}) = \prod_{k=1}^{K} \frac{\lambda_i(\tilde{\mathbf{F}}_k^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k\tilde{\mathbf{F}}_k)}{1 + \lambda_i(\tilde{\mathbf{F}}_k^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k\tilde{\mathbf{F}}_k)} \\
& \boldsymbol{\Theta} = \boldsymbol{A}_1^{\mathrm{H}}\cdots\boldsymbol{A}_K^{\mathrm{H}}\boldsymbol{A}_K\cdots\boldsymbol{A}_1 \\
& \boldsymbol{A}_k = (\boldsymbol{\mathcal{H}}_k\tilde{\mathbf{F}}_k\tilde{\mathbf{F}}_k^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k^{\mathrm{H}} + \mathbf{I}_{M_k})^{-1/2}\boldsymbol{\mathcal{H}}_k\tilde{\mathbf{F}}_k, \\
& \mathbf{U}_{\boldsymbol{A}_k} = \mathbf{V}_{\boldsymbol{A}_{k-1}}, \quad k = 2,\cdots,K \\
& \mathrm{Tr}(\tilde{\mathbf{F}}_k\tilde{\mathbf{F}}_k^{\mathrm{H}}) = P_k,
\end{aligned} \tag{41}
$$

where $\mathbf{U}_{\boldsymbol{A}_k}$ and $\mathbf{V}_{\boldsymbol{A}_k}$ are defined based on the singular value decomposition $\boldsymbol{A}_k = \mathbf{U}_{\boldsymbol{A}_k}\boldsymbol{\Lambda}_{\boldsymbol{A}_k}\mathbf{V}_{\boldsymbol{A}_k}^{\mathrm{H}}$ with $\boldsymbol{\Lambda}_{\boldsymbol{A}_k} \searrow$.

Structure of optimal $\tilde{\mathbf{F}}_k$: For the optimal $\tilde{\mathbf{F}}_k$, denoted as $\tilde{\mathbf{F}}_{k,\mathrm{opt}}$, based on the following singular value decompositions

$$
\boldsymbol{\mathcal{H}}_k\tilde{\mathbf{F}}_{k,\mathrm{opt}} = \mathbf{U}_{M_k}\boldsymbol{\Lambda}_{M_k}\mathbf{V}_{M_k}^{\mathrm{H}} \quad \text{with} \quad \boldsymbol{\Lambda}_{M_k} \searrow \tag{42}
$$

$$
\boldsymbol{\mathcal{H}}_k = \mathbf{U}_{\boldsymbol{\mathcal{H}}_k}\boldsymbol{\Lambda}_{\boldsymbol{\mathcal{H}}_k}\mathbf{V}_{\boldsymbol{\mathcal{H}}_k}^{\mathrm{H}} \quad \text{with} \quad \boldsymbol{\Lambda}_{\boldsymbol{\mathcal{H}}_k} \searrow, \tag{43}
$$

we can construct a matrix $\bar{\mathbf{F}}_k$

$$
\bar{\mathbf{F}}_k = \mathbf{V}_{\boldsymbol{\mathcal{H}}_k}\boldsymbol{\Lambda}_{\mathbf{X}_k}\mathbf{V}_{M_k}^{\mathrm{H}} \tag{44}
$$

where $\boldsymbol{\Lambda}_{\mathbf{X}_k}$ is an unknown diagonal matrix with the same rank as $\boldsymbol{\Lambda}_{M_k}$ and $\boldsymbol{\Lambda}_{\boldsymbol{\mathcal{H}}_k}\boldsymbol{\Lambda}_{\mathbf{X}_k}/b = \boldsymbol{\Lambda}_{M_k}$, and the scalar b is chosen to make $\mathrm{Tr}(\bar{\mathbf{F}}_k\bar{\mathbf{F}}_k^{\mathrm{H}}) = P_k$ hold.

Because $\boldsymbol{\mathcal{H}}_k$ is independent of unknown variable, using Lemma 12 in [11], we have

$$
\bar{\mathbf{F}}_k^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k\bar{\mathbf{F}}_k \succeq \tilde{\mathbf{F}}_{k,\mathrm{opt}}^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k\tilde{\mathbf{F}}_{k,\mathrm{opt}}. \tag{45}
$$

Taking eigenvalues of both sides, we have $\lambda_i(\bar{\mathbf{F}}_k^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k\bar{\mathbf{F}}_k) \geq \lambda_i(\tilde{\mathbf{F}}_{k,\mathrm{opt}}^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k\tilde{\mathbf{F}}_{k,\mathrm{opt}})$ [14]. Since $\gamma_i(\boldsymbol{\Theta})$ is an increasing function of $\lambda_i(\tilde{\mathbf{F}}_k^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k\tilde{\mathbf{F}}_k)$, we directly have that $\boldsymbol{\gamma}(\boldsymbol{\Theta}|_{\tilde{\mathbf{F}}_k = \bar{\mathbf{F}}_k}) \geq \boldsymbol{\gamma}(\boldsymbol{\Theta}|_{\tilde{\mathbf{F}}_k = \tilde{\mathbf{F}}_{k,\mathrm{opt}}})$. Furthermore, as the objective function $g(\bullet)$ of (41) is a decreasing function, we finally have $g[\boldsymbol{\gamma}(\boldsymbol{\Theta}|_{\tilde{\mathbf{F}}_k = \bar{\mathbf{F}}_k})] \leq g[\boldsymbol{\gamma}(\boldsymbol{\Theta}|_{\tilde{\mathbf{F}}_k = \tilde{\mathbf{F}}_{k,\mathrm{opt}}})]$. Because $\tilde{\mathbf{F}}_{k,\mathrm{opt}}$ is the optimal solution, $\tilde{\mathbf{F}}_{k,\mathrm{opt}}$ must be in the form of $\bar{\mathbf{F}}_k$. Therefore, the structure of optimal $\tilde{\mathbf{F}}_k$ is given by (44) $\tilde{\mathbf{F}}_{k,\mathrm{opt}} = \mathbf{V}_{\boldsymbol{\mathcal{H}}_k}\boldsymbol{\Lambda}_{\mathbf{X}_k}\mathbf{V}_{M_k}^{\mathrm{H}}$.

Notice that $\mathbf{V}_{M_k}$ is still unknown. Substituting $\tilde{\mathbf{F}}_{k,\mathrm{opt}}$ into the definition of $\boldsymbol{A}_k$ in (41), we have $\mathbf{V}_{M_k} = \mathbf{Q}_{k-1}\mathbf{U}_{\boldsymbol{\mathcal{H}}_{k-1}}$. It can be seen that the value of $\mathbf{Q}_k$ does not affect the value of $\boldsymbol{\Theta}$ in optimization problem (41). Without loss of generality, it is set as $\mathbf{Q}_{k,\mathrm{opt}} = \mathbf{I}_{M_k}$. As a result, $\mathbf{V}_{M_k} =$

$\mathbf{U}_{\boldsymbol{\mathcal{H}}_{k-1}}$ and $\mathbf{V}_{M_1} = \mathbf{U}_{\boldsymbol{\Omega}}$. For the considered system model, $\mathrm{Rank}(\boldsymbol{\Lambda}_{\mathbf{X}_k}) \leq N$, and thus only the $N \times N$ principal submatrix of $\boldsymbol{\Lambda}_{\mathbf{X}_k}$ can be nonzero, which is denoted as $\boldsymbol{\Lambda}_{\boldsymbol{\mathcal{F}}_k}$. In summary, $\tilde{\mathbf{F}}_{k,\mathrm{opt}}$ has the following structure

$$
\tilde{\mathbf{F}}_{k,\mathrm{opt}} = \mathbf{V}_{\boldsymbol{\mathcal{H}}_k,N}\boldsymbol{\Lambda}_{\boldsymbol{\mathcal{F}}_k}\mathbf{U}_{\boldsymbol{\mathcal{H}}_{k-1},N}^{\mathrm{H}} \tag{46}
$$

with $\mathbf{U}_{\boldsymbol{\mathcal{H}}_0} = \mathbf{U}_{\boldsymbol{\Omega}}$.

Structure of optimal $\mathbf{F}_k$: Based the relationship between $\mathbf{F}_k$ and $\tilde{\mathbf{F}}_k$ given in (38),

$$
\mathbf{F}_{k,\mathrm{opt}} = \sqrt{\eta_{f_k}}(\alpha_k P_k\boldsymbol{\Psi}_k + \sigma_{n_k}^2\mathbf{I}_{N_k})^{-1/2}\tilde{\mathbf{F}}_{k,\mathrm{opt}}. \tag{47}
$$

Putting the structure of $\mathbf{F}_{k,\mathrm{opt}}$ in (47) into η_{f_k} in (20), and η_{f_k} can be solved to be

$$
\begin{aligned}
\eta_{f_k} = \sigma_{n_k}^2/\{1 - &\alpha_k\mathrm{Tr}[\mathbf{V}_{\boldsymbol{\mathcal{H}}_k,N}^{\mathrm{H}}(\alpha_k P_k\boldsymbol{\Psi}_k + \sigma_{n_k}^2\mathbf{I}_{N_k})^{-1/2}\boldsymbol{\Psi}_k \\
&\times (\alpha_k P_k\boldsymbol{\Psi}_k + \sigma_{n_k}^2\mathbf{I}_{N_k})^{-1/2}\mathbf{V}_{\boldsymbol{\mathcal{H}}_k,N}\boldsymbol{\Lambda}_{\boldsymbol{\mathcal{F}}_k}^2]\}.
\end{aligned} \tag{48}
$$

As in (48) η_{f_k} is a function of $\boldsymbol{\Lambda}_{\boldsymbol{\mathcal{F}}_k}$, it is denoted as $\eta_{f_k} = \xi_k(\boldsymbol{\Lambda}_{\boldsymbol{\mathcal{F}}_k})$.

REFERENCES

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

[2] 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.

[3] R. Mo and Y. Chew, "Precoder design for non-regenerative MIMO relay systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 10, pp. 5041–5049, Oct. 2009.

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

[5] Y. Rong and Y. Hua, "Optimality of diagonalization of multi-hop MIMO relays," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 6068–6077, Dec. 2009.

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

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

[8] C. Xing, Z. Fei, Y.-C. Wu, S. Ma, and J. Kuang, "Robust transceiver design for AF MIMO relaying systems with column correlations," ICSPCC 2011, Xi'an China, Sep. 2011.

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

[10] M. Ding and S. D. Blostein, "Maximum mutual information design for MIMO systems with imperfect channel knowledge," *IEEE Trans. Infor. Theory*, vol. 56, no. 10, pp.4793–4801, Oct. 2010.

[11] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. on Signal Process.*, vol. 51, no. 9, pp. 2381–2401, Sep. 2003.

[12] F. Gao, R. Zhang, and Y.-C. Liang, "Channel estimation for OFDM modulated two-way relay networks," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4443–4455, Nov. 2009.

[13] B. Hassibi and B. M. Hochwald, "How much training is needed in mutiple-antenna wireless links?" *Trans. Infor. Theory,* vol. 49, no. 4, pp951–963, Apr. 2003.

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

[15] W. Zhang, U. Mitra, and M. Chiang, "Optimization of amplify-and-forward multicarrier two-hop transmission," *IEEE Trans. Commun.*, vol. 59, no. 5, pp. 1434–1445, May 2011.