Title	Robust Tomlinson-Harashima precoding for non-regenerative multi-antenna relaying systems
Author(s)	Xing, C; Xia, M; Gao, F; 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. 753-758
Issued Date	2012
URL	http://hdl.handle.net/10722/165233
Rights	IEEE Wireless Communications and Networking Conference Proceedings. Copyright © IEEE Communications Society.

Robust Tomlinson-Harashima Precoding for Non-Regenerative Multi-Antenna Relaying Systems

Chengwen Xing[†], Minghua Xia[§], Feifei Gao[‡,*] and Yik-Chung Wu[£]

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

[§]Division of Physical Sciences and Engineering, King Abdullah University of Science and Technology, Arabia

[‡]National Mobile Communications Research Laboratory, Southeast University, Nanjing, China

[*]State Key Laboratory of Intelligent Technology and Systems, Tsinghua National Laboratory for Information Science and Technology, Department of Automation, Tsinghua University, Beijing, 100084, China

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

xingchengwen@gmail.com, minghua.xia@kaust.edu.sa, feifeigao@ieee.org, ycwu@eee.hku.hk.

Abstract—In this paper, we consider the robust transceiver design with Tomlinson-Harashima precoding (THP) for multi-hop amplify-and-forward (AF) multiple-input multiple-output (MIMO) relaying systems. THP is adopted at the source to mitigate the spatial inter-symbol interference and then a joint Bayesian robust design of THP at source, linear forwarding matrices at relays and linear equalizer at destination is proposed. Based on the elegant characteristics of multiplicative convexity and matrix-monotone functions, the optimal structure of the nonlinear transceiver is first derived. Based on the derived structure, the optimization problem is greatly simplified and can be efficiently solved. Finally, the performance advantage of the proposed robust design is assessed by simulation results.

I. INTRODUCTION

Transceiver design for amplify-and-forward (AF) multiple-input multiple-output (MIMO) relaying systems attracted a lot of attention recently, as it has a great potential to enhance the communication range of a simple point-to-point system, while providing spatial diversity and multiplexing gains. This system has been considered to be adopted in the emerging wireless systems, such as LTE-advanced, WINNER project, etc.

Linear transceiver design for dual-hop AF MIMO relaying systems has been extensively studied in [1]–[4]. For multiple-antenna systems, it is well-known that nonlinear transceivers have much better performance than their linear counterparts [5], [6]. Recently, nonlinear transceiver design for AF MIMO relaying systems assuming perfect CSI appears in [7]. There are two kinds of nonlinear transceiver design: decision-feedback equalization (DFE) based design and Tomlinson-Harashima precoding (THP) based design. In fact, there exits a duality between these two designs [7]. However, as THP is performed at transmitter, it is free of error propagation compared to DFE based one. THP is the

transmitter counterpart of the vertical BELL-Labs Layered Space-Time (V-BLAST) system. THP can effectively mitigate intersymbol interference or multi-user interference, and is also widely used as one-dimensional dirty paper coding (DPC). Due to its nonlinear nature, unfortunately, THP is more sensitive to channel estimation errors than its linear counterpart. In the presence of channel estimation errors, the performance of THP would degrade severely. Therefore, robust nonlinear transceiver design is a promising way to mitigate such problem.

In this paper, we consider a general multi-hop AF MIMO relaying system. The THP at the source, linear forwarding matrices at multiple relays and linear destination equalizer matrix are jointly optimized under channel estimation errors at all terminals. As many design objectives of THP can be considered as a multiplicatively Schur-convex or multiplicatively Schur-concave function, in this work, a unified optimization problem is investigated whose objective functions are multiplicatively Schur-convex/concave. With novel applications of results in multiplicative convexity and matrix-monotone functions, the optimal diagonal structure of the transceiver is derived. Then the transceiver design is then significantly simplified and then iterative water-filling alike solutions are adopted to solve for the remaining unknown variables. The performance advantage of the proposed robust design is assessed by simulations and is shown to perform much better than the corresponding non-robust design.

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 an $M \times M$ 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. The symbol $\mathbb{E}\{\bullet\}$ represents the statistical expectation. For two Hermitian matrices, $\mathbf{C} \succeq \mathbf{D}$ means that $\mathbf{C} - \mathbf{D}$ is a positive semi-definite matrix. The $(n, m)^{\mathrm{th}}$ entry of a matrix $\mathbf{Z}$ is denoted as $[\mathbf{Z}]_{n,m}$.

II. SIGNAL MODEL AND PROBLEM FORMULATION

A. Signal Model

In this paper, a K-hop amplify-and-forward MIMO relaying system is investigated, in which there is one source, one destination and $K - 1$ relays. The source is equipped with N_1 transmit antennas. The kth relay has M_k receive antennas and N_{k+1} transmit antennas. The destination is equipped with M_K receive antennas. At the source, at each time slot, there is a $N \times 1$ vector $\mathbf{a} = [a_1, a_2, \cdots, a_N]^{\text{T}}$ to be transmitted. Specifically, the data symbols are chosen from M-QAM constellation with the real and imaginary parts of a_k belong to the set $\mathcal{A} = \{\pm 1, \pm 3, \cdots, \pm(\sqrt{M} - 1)\}$.

At the transmitter, the data vector $\mathbf{a}$ is fed into the a precoding unit which consists of a $N \times N$ feedback matrix $\mathbf{B}$ and a nonlinear modulo operator $\text{MOD}_M(\bullet)$. The square matrix $\mathbf{B}$ is a strictly lower triangular matrix which allows data precoding in a recursive fashion and the $\text{MOD}_M(\bullet)$ is defined as

$$\text{MOD}_M(x) = x - 2\sqrt{M}\left[\left\lfloor \frac{\text{Re}(x)}{2\sqrt{M}} + \frac{1}{2} \right\rfloor \right.$$
$$\left. + \sqrt{-1}\left\lfloor \frac{\text{Im}(x)}{2\sqrt{M}} + \frac{1}{2} \right\rfloor \right], \quad (1)$$

where the symbol $\lfloor z \rfloor$ denotes the largest integer not exceeding z. The nonlinear modulo operator reduces the output signals into a square region $[-\sqrt{M}, \sqrt{M}) \times [-\sqrt{M}, \sqrt{M})$.

Generally speaking, nonlinear operation is more complicated to be analyzed than linear operation. To simplify the following analysis, the nonlinear precoder can be interpreted as the following linear operation as

$$b_k = a_k - \sum_{l=1}^{k-1} [\mathbf{B}]_{k,l} b_l + d_k \quad (2)$$

where $d_k = 2\sqrt{M}I_k$ and I_k is a complex number whose real and imaginary components are both integer. While we do not need to know the exact value of d_k, it has the effect of reducing b_k into the square region $[-\sqrt{M}, \sqrt{M}) \times [-\sqrt{M}, \sqrt{M})$. The previous equation can be written into a compact form as

$$\mathbf{b} = \underbrace{(\mathbf{B} + \mathbf{I}_N)^{-1}}_{\triangleq \mathbf{C}} \underbrace{(\mathbf{a} + \mathbf{d})}_{\triangleq \mathbf{s}} \quad (3)$$

where $\mathbf{b} \triangleq [b_1, \cdots, b_N]^{\text{T}}$, $\mathbf{d} \triangleq [d_1, \cdots, d_N]^{\text{T}}$, and $\mathbf{C}$ is a lower triangular matrix with unit diagonal elements, i.e., $[\mathbf{C}]_{k,l} = 0$ for $k < l$ and $[\mathbf{C}]_{k,k} = 1$.

After the nonlinear operation, the vector $\mathbf{b}$ is multiplied with a precoder matrix $\mathbf{P}_1$ under a transmit power constraint $\text{Tr}(\mathbf{P}_1 \mathbf{R_b} \mathbf{P}_1^{\text{H}}) \leq P_1$ where P_1 is the maximum transmit power at the source. As the elements of $\mathbf{a}$ are independent and identically distributed (i.i.d.) over the constellation, $\mathbf{b}$ can also be considered as i.i.d. [10], i.e.,

$$\mathbf{R_b} = 2(M-1)/3\mathbf{I}_N \triangleq \sigma_b^2 \mathbf{I}_N. \quad (4)$$

The received signal $\mathbf{x}_1$ at the first relay is formulated as $\mathbf{x}_1 = \mathbf{H}_1 \mathbf{P}_1 \mathbf{b} + \mathbf{n}_1$ where $\mathbf{H}_1$ is the channel between the source and the first relay and $\mathbf{n}_1$ is additive Gaussian noise with mean zero and covariance matrix $\mathbf{R}_{\mathbf{n}_1} = \sigma_{n_1}^2 \mathbf{I}_{M_1}$.

Similarly, at the kth relay the received signal is

$$\mathbf{x}_k = \mathbf{H}_k \mathbf{P}_k \mathbf{x}_{k-1} + \mathbf{n}_k \quad (5)$$

with $\mathbf{H}_k$ and $\mathbf{n}_k$ are the channel and additive noise at the kth hop, respectively. The covariance matrix of $\mathbf{n}_k$ is denoted as $\mathbf{R}_{\mathbf{n}_k} = \sigma_{n_k}^2 \mathbf{I}_{M_k}$. Finally, for a K-hop AF MIMO relaying system, the received signal at the destination is

$$\mathbf{y} = \left[\prod_{k=1}^{K} \mathbf{H}_k \mathbf{P}_k \right] \mathbf{b} + \sum_{k=1}^{K-1}\left\{ \left[\prod_{l=k+1}^{K} \mathbf{H}_l \mathbf{P}_l \right] \mathbf{n}_k \right\} + \mathbf{n}_K, \quad (6)$$

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 [3].

In practice, the channels $\mathbf{H}_k$ are estimated and channel estimation errors are inevitable. Therefore, the channel $\mathbf{H}_k$ can be expressed as

$$\mathbf{H}_k = \bar{\mathbf{H}}_k + \Delta\mathbf{H}_k, \quad (7)$$

where $\bar{\mathbf{H}}_k$ is the estimated channels, and $\Delta\mathbf{H}_k$ is the corresponding channel estimation errors whose elements are zero mean Gaussian random variables. Furthermore, the $M_k \times N_k$ matrix $\Delta\mathbf{H}_k$ can be decomposed using the widely used Kronecker model [8] as $\Delta\mathbf{H}_k = \mathbf{\Sigma}_k^{1/2}\mathbf{H}_{W,k}\mathbf{\Psi}_k^{1/2}$, where the elements of the $M_k \times N_k$ matrix $\mathbf{H}_{W,k}$ are i.i.d. Gaussian random variables with zero mean and unit variance. The specific formulas of $\mathbf{\Sigma}_k$ and $\mathbf{\Psi}_k$ are determined by the training sequences and channel estimators [8], [9]

B. Problem Formulation

At the destination, a linear equalizer $\mathbf{G}$ is adopted and is followed by a modulo operator. Notice that the effect of $\mathbf{d}$ will be perfectly removed by modulo operator at the destination and estimating $\mathbf{a}$ is equivalent to estimating $\mathbf{s}$ [6]. Thus at the destination, a linear equalizer $\mathbf{G}$ is used to detect the data vector $\mathbf{s}$. The MSE matrix of the data vector is defined as $\mathbb{E}\{(\mathbf{G}\mathbf{y} - \mathbf{s})(\mathbf{G}\mathbf{y} - \mathbf{s})^{\text{H}}\}$ [6], [10], where the expectation is taken with respect to random data, channel estimation errors, and noise. Following a similar derivation to that in [8], it can be shown that

$$\mathbf{\Phi}(\mathbf{G}, \mathbf{P}_k, \mathbf{C})$$
$$= \mathbb{E}\{(\mathbf{G}\mathbf{y} - \mathbf{C}\mathbf{b})(\mathbf{G}\mathbf{y} - \mathbf{C}\mathbf{b})^{\text{H}}\}$$
$$= \mathbf{G}[\bar{\mathbf{H}}_K \mathbf{P}_K \mathbf{R}_{\mathbf{x}_{K-1}} \mathbf{P}_K^{\text{H}} \bar{\mathbf{H}}_K^{\text{H}} + \text{Tr}(\mathbf{P}_K \mathbf{R}_{\mathbf{x}_{K-1}} \mathbf{P}_K^{\text{H}} \mathbf{\Psi}_K)\mathbf{\Sigma}_K$$
$$+ \mathbf{R}_{n_K}]\mathbf{G}^{\text{H}} - \sigma_b^2 \mathbf{G} \prod_{k=1}^{K}\left(\bar{\mathbf{H}}_k \mathbf{P}_k\right)\mathbf{C}^{\text{H}} + \sigma_b^2 \mathbf{C}\mathbf{C}^{\text{H}}$$
$$- \sigma_b^2 \left[\mathbf{G} \prod_{k=1}^{K}\left(\bar{\mathbf{H}}_k \mathbf{P}_k\right)\mathbf{C}^{\text{H}}\right]^{\text{H}} \quad (8)$$

where matrices $\mathbf{R}_{\mathbf{x}_k}$ is defined as

$$\mathbf{R}_{\mathbf{x}_k} \triangleq \mathbb{E}\{\mathbf{x}_k\mathbf{x}_k^{\text{H}}\} = \bar{\mathbf{H}}_k \mathbf{P}_k \mathbf{R}_{\mathbf{x}_{k-1}} \mathbf{P}_k^{\text{H}} \bar{\mathbf{H}}_k^{\text{H}} + \mathbf{R}_{\mathbf{n}_k}$$
$$+ \text{Tr}(\mathbf{P}_k \mathbf{R}_{\mathbf{x}_{k-1}} \mathbf{P}_k^{\text{H}} \mathbf{\Psi}_k)\mathbf{\Sigma}_k. \quad (9)$$

It is obvious that $\mathbf{R}_{\mathbf{x}_k}$ is the covariance matrix of the received signal at the relay. Notice that $\mathbf{R}_{\mathbf{x}_0} = \mathbf{R}_{\mathbf{b}} = \sigma_b^2 \mathbf{I}_N$.

For MIMO transceiver design, a wide range of objective functions can be expressed as a function of the diagonal elements of the MSE matrix. For example, for sum MSE minimization, the objective function is $f([\mathrm{MSE}_1, \cdots, \mathrm{MSE}_N]^{\mathrm{T}}) = \sum_{n=1}^{N} \mathrm{MSE}_n$, where $\mathrm{MSE}_n = [\mathbf{\Phi}(\mathbf{G}, \mathbf{P}_k, \mathbf{C})]_{n,n}$. For product MSE minimization, the objective function is $f([\mathrm{MSE}_1, \cdots, \mathrm{MSE}_N]^{\mathrm{T}}) = \prod_{n=1}^{N} \mathrm{MSE}_n$. Furthermore, worst-case MSE minimization corresponds to minimizing the objective function given as $f([\mathrm{MSE}_1, \cdots, \mathrm{MSE}_N]^{\mathrm{T}}) = \max_{n=1,2,\cdots,N}\{\mathrm{MSE}_n\}$ [6]. On the other hand, weighted geometric mean MSE minimization corresponds to minimizing the following objective function $f([\mathrm{MSE}_1, \cdots, \mathrm{MSE}_N]^{\mathrm{T}}) = \prod_{n=1}^{N} \mathrm{MSE}_n^{w_n}$ with $w_1 \geq w_2 \cdots \geq w_N \geq 0$. Therefore, a unified transceiver design optimization problem can be formulated as

$$\min_{\mathbf{G}, \mathbf{P}_k, \mathbf{C}} \quad f([\mathrm{MSE}_1, \cdots, \mathrm{MSE}_N]^{\mathrm{T}})$$
$$\mathrm{s.t.} \quad \mathrm{MSE}_n = [\mathbf{\Phi}(\mathbf{G}, \mathbf{P}_k, \mathbf{C})]_{n,n}$$
$$\mathrm{Tr}(\mathbf{P}_k \mathbf{R}_{\mathbf{x}_{k-1}} \mathbf{P}_k^{\mathrm{H}}) \leq P_k, \quad k = 1, \cdots, K \quad (10)$$

where the matrix $\mathbf{C}$ is a lower triangular matrix with unit diagonal elements.

In general, the objective function $f(\bullet)$ possesses two important properties:

(1) $f(\bullet)$ is an increasing real-valued matrix function $\mathbb{C}^N \to \mathbb{R}$, i.e., for two vectors $\mathbf{u} = [u_1, u_2, \cdots, u_N]^{\mathrm{T}}$ and $\mathbf{v} = [v_1, v_2, \cdots, v_N]^{\mathrm{T}}$, when $u_n \geq v_n$, we have $f(\mathbf{u}) \geq f(\mathbf{v})$. This property is natural in transceiver design. This fact is reflected in $f(\bullet)$ being an increasing function.
(2) $f(\bullet)$ is multiplicatively Schur-convex or concave.

The detailed discussions are given in [11]. In the following, for notational convenience, multiplicatively Schur-convex/concave is referred to as M-Schur-convex/concave.

III. OPTIMAL DESIGN OF $\mathbf{G}$ AND $\mathbf{C}$

The linear minimum mean-square-error (LMMSE) equalizer is obtained by setting the differentiation of the trace of (8) with respect to $\mathbf{G}^*$ to be zero, and we have

$$\mathbf{G}_{\mathrm{LMMSE}}$$
$$= \sigma_b^2 \left[\prod_{k=1}^{K} (\bar{\mathbf{H}}_k \mathbf{P}_k) \mathbf{C}^{\mathrm{H}}\right]^{\mathrm{H}} [\bar{\mathbf{H}}_K \mathbf{P}_K \mathbf{R}_{\mathbf{x}_{K-1}} \mathbf{P}_K^{\mathrm{H}} \bar{\mathbf{H}}_K^{\mathrm{H}}$$
$$+ \mathrm{Tr}(\mathbf{P}_K \mathbf{R}_{\mathbf{x}_{K-1}} \mathbf{P}_K^{\mathrm{H}} \mathbf{\Psi}_K) \mathbf{\Sigma}_K + \mathbf{R}_{\mathbf{n}_K}]^{-1}. \quad (11)$$

In terms of MSE, LMMSE estimator is a dominated estimator in linear estimators, i.e.,

$$\mathbf{\Phi}(\mathbf{G}_{\mathrm{LMMSE}}, \mathbf{P}_k, \mathbf{C}) \preceq \mathbf{\Phi}(\mathbf{G}, \mathbf{P}_k, \mathbf{C}) \quad (12)$$

which implies $[\mathbf{\Phi}(\mathbf{G}_{\mathrm{LMMSE}}, \mathbf{P}_k, \mathbf{C})]_{n,n} \leq [\mathbf{\Phi}(\mathbf{G}, \mathbf{P}_k, \mathbf{C})]_{n,n}$. As $f(\bullet)$ is an increasing function, and there is no constraint on $\mathbf{G}$ in (10), the optimal linear equalizer is LMMSE equalizer, i.e., $\mathbf{G}_{\mathrm{opt}} = \mathbf{G}_{\mathrm{LMMSE}}$.

Substituting the optimal equalizer (11) into the MSE formulation (8), the MSE matrix is rewritten as

$$\mathbf{\Phi}_{\mathrm{MSE}}(\mathbf{P}_k, \mathbf{C}) =$$
$$\sigma_b^2 \mathbf{C} \Big(\mathbf{I}_N - \sigma_b^2 \left[\prod_{k=1}^{K} (\bar{\mathbf{H}}_k \mathbf{P}_k) \mathbf{C}^{\mathrm{H}}\right]^{\mathrm{H}} [\bar{\mathbf{H}}_K \mathbf{P}_K \mathbf{R}_{\mathbf{x}_{K-1}} \mathbf{P}_K^{\mathrm{H}} \bar{\mathbf{H}}_K^{\mathrm{H}}$$
$$+ \mathrm{Tr}(\mathbf{P}_K \mathbf{R}_{\mathbf{x}_{K-1}} \mathbf{P}_K^{\mathrm{H}} \mathbf{\Psi}_K) \mathbf{\Sigma}_K + \mathbf{R}_{\mathbf{n}_K}]^{-1} \left[\prod_{k=1}^{K} (\bar{\mathbf{H}}_k \mathbf{P}_k) \mathbf{C}^{\mathrm{H}}\right] \Big)$$
$$\times \mathbf{C}^{\mathrm{H}} \quad (13)$$

From the definition of $\mathbf{R}_{\mathbf{x}_k}$ in (9), it is noticed that $\mathbf{R}_{\mathbf{x}_k}$ is a function of $\mathbf{P}_l$ with $l \leq k$. In other words, the constraints are coupled with each other. In order to simplify the analysis, we define the following new variables

$$\mathbf{F}_1 = \mathbf{P}_1 \mathbf{R}_{\mathbf{b}}^{1/2} \mathbf{Q}_0^{\mathrm{H}} \quad (14)$$
$$\mathbf{F}_k = \mathbf{P}_k \mathbf{K}_{\mathbf{F}_{k-1}}^{1/2} (\mathbf{K}_{\mathbf{F}_{k-1}}^{-1/2} \bar{\mathbf{H}}_{k-1} \mathbf{F}_{k-1} \mathbf{F}_{k-1}^{\mathrm{H}} \bar{\mathbf{H}}_{k-1}^{\mathrm{H}} \mathbf{K}_{\mathbf{F}_{k-1}}^{-1/2}$$
$$+ \mathbf{I}_{M_{k-1}})^{1/2} \mathbf{Q}_{k-1}^{\mathrm{H}} \quad (15)$$

where $\mathbf{K}_{\mathbf{F}_k}$ is defined as

$$\mathbf{K}_{\mathbf{F}_k} \triangleq \mathrm{Tr}(\mathbf{F}_k \mathbf{F}_k^{\mathrm{H}} \mathbf{\Psi}_k) \mathbf{\Sigma}_k + \sigma_{n_k}^2 \mathbf{I}_{M_k}, \quad (16)$$

and the matrix $\mathbf{Q}_k$ is an additional unknown unitary matrix. Based on the definition of $\mathbf{F}_k$ in (14) and (15), it is easy to show that $\mathbf{F}_k \mathbf{F}_k^{\mathrm{H}} = \mathbf{P}_k \mathbf{R}_{\mathbf{x}_{k-1}} \mathbf{P}_k^{\mathrm{H}}$ and thus the power constraints becomes independent of each other

$$\mathrm{Tr}(\mathbf{P}_k \mathbf{R}_{\mathbf{x}_{k-1}} \mathbf{P}_k^{\mathrm{H}}) = \mathrm{Tr}(\mathbf{F}_k \mathbf{F}_k^{\mathrm{H}}) \leq P_k. \quad (17)$$

Meanwhile, using (14) and (15), defining

$$\mathbf{M}_k = (\mathbf{K}_{\mathbf{F}_k}^{-1/2} \bar{\mathbf{H}}_k \mathbf{F}_k \mathbf{F}_k^{\mathrm{H}} \bar{\mathbf{H}}_k^{\mathrm{H}} \mathbf{K}_{\mathbf{F}_k}^{-1/2} + \mathbf{I}_{M_k})^{-1/2}$$
$$\times \mathbf{K}_{\mathbf{F}_k}^{-1/2} \bar{\mathbf{H}}_k \mathbf{F}_k, \quad (18)$$

the optimization problem (10) is simplified as

$$\min_{\mathbf{F}_k, \mathbf{Q}_k, \mathbf{C}} \quad f([\mathrm{MSE}_1, \cdots, \mathrm{MSE}_N]^{\mathrm{T}})$$
$$\mathrm{s.t.} \quad \mathrm{MSE}_n = \sigma_b^2 \left[\mathbf{C}(\mathbf{I}_N - \mathbf{Q}_0^{\mathrm{H}} \mathbf{\Theta} \mathbf{Q}_0) \mathbf{C}^{\mathrm{H}}\right]_{n,n}$$
$$\mathbf{\Theta} = \mathbf{M}_1^{\mathrm{H}} \mathbf{Q}_1^{\mathrm{H}} \cdots \mathbf{M}_K^{\mathrm{H}} \mathbf{Q}_K^{\mathrm{H}} \mathbf{Q}_K \mathbf{M}_K \cdots \mathbf{Q}_1 \mathbf{M}_1$$
$$\mathrm{Tr}(\mathbf{F}_k \mathbf{F}_k^{\mathrm{H}}) \leq P_k, \quad \mathbf{Q}_k^{\mathrm{H}} \mathbf{Q}_k = \mathbf{I}_{M_{k-1}}. \quad (19)$$

Notice that the largest singular value of $\mathbf{M}_k$ is smaller than one. Therefore, the largest eigenvalue of $\mathbf{\Theta}$ is smaller than one [12] and then $\mathbf{I}_N - \mathbf{Q}_0^{\mathrm{H}} \mathbf{\Theta} \mathbf{Q}_0$ is a positive definite matrix. With the Cholesky factorization

$$(\mathbf{I}_N - \mathbf{Q}_0^{\mathrm{H}} \mathbf{\Theta} \mathbf{Q}_0) \sigma_b^2 = \mathbf{L} \mathbf{L}^{\mathrm{H}} \quad (20)$$

where $\mathbf{L}$ is a lower triangular matrix. It can be derived that the optimal $\mathbf{C}$ equals to

$$\mathbf{C}_{\mathrm{opt}} = \mathbf{D} \mathbf{L}^{-1}. \quad (21)$$

where $\mathbf{D}$ is a diagonal matrix defined as

$$\mathbf{D} = \mathrm{diag}\{[\mathbf{L}_{1,1}, \cdots, \mathbf{L}_{N,N}]^{\mathrm{T}}\}. \quad (22)$$

As a result $\text{MSE}_n = [\mathbf{L}]_{n,n}^2$, and the optimization problem for robust transceiver design is formulated as

$$\min_{\mathbf{F}_k, \mathbf{Q}_k} \quad f\left([[\mathbf{L}]_{1,1}^2, \cdots, [\mathbf{L}]_{N,N}^2]^\text{T}\right)$$

$$\text{s.t.} \quad \sigma_b^2(\mathbf{I}_N - \mathbf{Q}_0^\text{H}\boldsymbol{\Theta}\mathbf{Q}_0) = \mathbf{L}\mathbf{L}^\text{H}$$

$$\boldsymbol{\Theta} = \mathbf{M}_1^\text{H}\mathbf{Q}_1^\text{H}\cdots\mathbf{M}_K^\text{H}\mathbf{Q}_K^\text{H}\mathbf{Q}_K\mathbf{M}_K\cdots\mathbf{Q}_1\mathbf{M}_1$$

$$\text{Tr}(\mathbf{F}_k\mathbf{F}_k^\text{H}) \le P_k, \quad \mathbf{Q}_0^\text{H}\mathbf{Q}_k = \mathbf{I}_{M_k}. \quad (23)$$

IV. OPTIMIZATION PROBLEM REFORMULATION FOR $\mathbf{F}_k$

A. Optimal Solution of $\mathbf{Q}_0$

Because the objective function of the optimization problem (23) is M-Schur-convex or M-Schur-concave (The details can be found in [11]). In the following, we will discuss the two cases separately.

M-Schur-convex:

Taking the determinant on both sides of (20), we have

$$|\sigma_b^2(\mathbf{I}_N - \mathbf{Q}_0^\text{H}\boldsymbol{\Theta}\mathbf{Q}_0)| = \prod_{n=1}^{N}[\mathbf{L}]_{n,n}^2 = \sigma_b^{2N}\prod_{n=1}^{N}(1 - \lambda_n(\boldsymbol{\Theta})) \quad (24)$$

where $\lambda_n(\boldsymbol{\Theta})$ is the nth largest eigenvalue of $\boldsymbol{\Theta}$. Based on (24), the following multiplicative majorization relationship can be established [11]

$$\sigma_b^2\left[\prod_{n=1}^{N}(1 - \lambda_n(\boldsymbol{\Theta}))\right]^{\frac{1}{N}} \otimes \mathbf{1}_N \prec_\times [[\mathbf{L}]_{1,1}^2, \cdots, [\mathbf{L}]_{N,N}^2]^\text{T}, \quad (25)$$

where the symbol $\otimes$ denotes the Kronecker product and $\mathbf{1}_N$ is a $N \times 1$ all-one vector. Moreover, $\mathbf{v} \prec_\times \mathbf{u}$ denotes $\mathbf{v}$ is multiplicatively majorized by $\mathbf{u}$ [11]. As $f(\bullet)$ being a M-Schur-convex function, (25) leads to

$$f\left([[\mathbf{L}]_{1,1}^2, \cdots, [\mathbf{L}]_{N,N}^2]^\text{T}\right)$$

$$\ge \underbrace{f\left(\sigma_b^2\left[\prod_{n=1}^{N}(1 - \lambda_n(\boldsymbol{\Theta}))\right]^{\frac{1}{N}} \otimes \mathbf{1}_N\right)}_{\triangleq g[\boldsymbol{\lambda}(\boldsymbol{\Theta})]}, \quad (26)$$

where $\boldsymbol{\lambda}(\boldsymbol{\Theta}) = [\lambda_1(\boldsymbol{\Theta}), \cdots, \lambda_N(\boldsymbol{\Theta})]^\text{T}$. The equality in (26) holds when $\prec_\times$ in (25) is replaced by equality, which means that $[\mathbf{L}]_{n,n}^2$ are identical for all n. Notice that from (20), we can write $\mathbf{L}\mathbf{L}^\text{H} = \sigma_b^2\mathbf{Q}_0^\text{H}(\mathbf{I}-\boldsymbol{\Theta})\mathbf{Q}_0$. Since $\mathbf{I}-\boldsymbol{\Theta}$ is positive definite, there always exists an unitary matrix $\mathbf{Q}_0$ which makes the Chrolesky factorization matrix of $\mathbf{Q}_0^\text{H}(\mathbf{I}-\boldsymbol{\Theta})\mathbf{Q}_0$ have identical diagonal elements [6].

M-Schur-concave:

From definition of $\mathbf{L}$ in (20) and based Weyl' theorem [6], we have [11]

$$[[\mathbf{L}]_{1,1}^2, \cdots, [\mathbf{L}]_{N,N}^2]^\text{T} \prec_\times \sigma_b^2[\mathbf{1}_N - \boldsymbol{\lambda}(\boldsymbol{\Theta})]. \quad (27)$$

Applying $f(\bullet)$ on both sides of (27), we have

$$f([[\mathbf{L}]_{1,1}^2, \cdots, [\mathbf{L}]_{N,N}^2]^\text{T}) \ge \underbrace{f(\sigma_b^2[\mathbf{1}_N - \boldsymbol{\lambda}(\boldsymbol{\Theta})])}_{\triangleq g[\boldsymbol{\lambda}(\boldsymbol{\Theta})]}. \quad (28)$$

The equality in (28) holds when $\prec_\times$ in (27) is replaced by equality, which means that $[\mathbf{L}]_{n,n}^2$ equals to $\sigma_b^2[1 - \lambda_n(\boldsymbol{\Theta})]$. On the other hand, taking eigenvalues on both sides of (20), we can obtain $\sigma_b^2[\mathbf{1}_N - \boldsymbol{\lambda}(\boldsymbol{\Theta})] = [\lambda_N(\mathbf{L}\mathbf{L}^\text{H}), \cdots, \lambda_1(\mathbf{L}\mathbf{L}^\text{H})]^\text{T}$. Therefore, $[[\mathbf{L}]_{1,1}^2, \cdots, [\mathbf{L}]_{N,N}^2]^\text{T} = [\lambda_N(\mathbf{L}\mathbf{L}^\text{H}), \cdots, \lambda_1(\mathbf{L}\mathbf{L}^\text{H})]^\text{T}$, which implies L is a diagonal matrix. With $\mathbf{L}$ being a diagonal matrix, $\mathbf{Q}_0^\text{H}\boldsymbol{\Theta}\mathbf{Q}_0$ is also a diagonal matrix. This can be satisfied if we take $\mathbf{Q}_0 = \mathbf{U}_\Theta$, where the unitary matrix $\mathbf{U}_\Theta$ is defined based on the eigendecomposition $\boldsymbol{\Theta} = \mathbf{U}_\Theta\boldsymbol{\Lambda}_\Theta\mathbf{U}_\Theta^\text{H}$ with the elements of $\boldsymbol{\Lambda}_\Theta$ arranged in decreasing order.

B. Problem Reformulation

Based on the given results of multiplicative majorization theory, the optimization problem (23) can be transformed into a much simpler one. Before presenting the result, two useful properties of the objective function $g(\bullet)$ are first derived based on the multiplicative majorization theory.

Property 1: The vector $\boldsymbol{\lambda}(\boldsymbol{\Theta})$ has the following weak multiplicative majorization relationship (Notice that $\mathbf{v} \prec_\Pi \mathbf{u}$ denotes $\mathbf{v}$ is weakly multiplicatively majorized by $\mathbf{u}$.) [11]

$$\boldsymbol{\lambda}(\boldsymbol{\Theta}) \prec_\Pi$$

$$\underbrace{[\gamma_1(\{\mathbf{F}_k\}_{k=1}^K), \gamma_2(\{\mathbf{F}_k\}_{k=1}^K), \cdots, \gamma_N(\{\mathbf{F}_k\}_{k=1}^K)]^\text{T}}_{\triangleq \boldsymbol{\gamma}(\{\mathbf{F}_k\}_{k=1}^K)}$$

$$\text{with } \gamma_n(\{\mathbf{F}_k\}_{k=1}^K) = \prod_{k=1}^{K}\frac{\lambda_n(\mathbf{F}_k^\text{H}\bar{\mathbf{H}}_k^\text{H}\mathbf{K}_{\mathbf{F}_k}^{-1}\bar{\mathbf{H}}_k\mathbf{F}_k)}{1 + \lambda_n(\mathbf{F}_k^\text{H}\bar{\mathbf{H}}_k^\text{H}\mathbf{K}_{\mathbf{F}_k}^{-1}\bar{\mathbf{H}}_k\mathbf{F}_k)}, \quad (29)$$

where the equality holds when

$$\mathbf{Q}_k = \mathbf{V}_{\mathbf{M}_{k+1}}\mathbf{U}_{\mathbf{M}_k}^\text{H}, \quad k = 1, \cdots, K-1 \quad (30)$$

where $\mathbf{U}_{\mathbf{M}_k}$ and $\mathbf{V}_{\mathbf{M}_k}$ are defined based on the singular value decomposition $\mathbf{M}_k = \mathbf{U}_{\mathbf{M}_k}\boldsymbol{\Lambda}_{\mathbf{M}_k}\mathbf{V}_{\mathbf{M}_k}^\text{H}$ with the diagonal elements of $\boldsymbol{\Lambda}_{\mathbf{M}_k}$ arranged in decreasing order. Notice that (30) does not cover $\mathbf{Q}_K$, but it can be any unitary matrix because it always appears in the form $\mathbf{Q}_K^\text{H}\mathbf{Q}_K$ and equals to an identity matrix in the objective function.

Proof: See Appendix A. ∎

Property 2: The objective function $g[\boldsymbol{\lambda}(\boldsymbol{\Theta})]$ in the both cases is a decreasing M-Schur-concave function with respective to $\boldsymbol{\lambda}(\boldsymbol{\Theta})$.

Based on **Properties 1** and **2**, the objective function has an achievable lower bound $g[\boldsymbol{\lambda}(\boldsymbol{\Theta})] \ge g[\boldsymbol{\gamma}(\{\mathbf{F}_k\}_{k=1}^K)]$ with equality achieved when (30) is satisfied [11]. When the lower bound is achieved, we have the following three additional observations:

(a) The constraints $\mathbf{Q}_k^\text{H}\mathbf{Q}_k = \mathbf{I}_{M_k}$ are automatically satisfied. (b) The objective function $g[\boldsymbol{\gamma}(\{\mathbf{F}_k\}_{k=1}^K)]$ is independent of $\mathbf{Q}_k$. (c) When $\mathbf{F}_k$'s are known, $\mathbf{Q}_k$'s can be directly computed using (30).

Applying these three observations into (23), we have the

reformulated optimization problem

$$\min_{\mathbf{F}_k} \quad g[\boldsymbol{\gamma}(\{\mathbf{F}_k\}_{k=1}^K)]$$

$$\text{s.t.} \quad \gamma_n(\{\mathbf{F}_k\}_{k=1}^K) = \prod_{k=1}^K \frac{\lambda_n(\mathbf{F}_k^H \bar{\mathbf{H}}_k^H \mathbf{K}_{\mathbf{F}_k}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_k)}{1 + \lambda_n(\mathbf{F}_k^H \bar{\mathbf{H}}_k^H \mathbf{K}_{\mathbf{F}_k}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_k)}$$

$$\text{Tr}(\mathbf{F}_k \mathbf{F}_k^H) \le P_k. \tag{31}$$

V. SOLUTION OF $\mathbf{F}_k$

In the following, we first derive the optimal structure of $\mathbf{F}_k$ and then present an algorithm to solve for the remaining unknown variables.

A. Optimal Structure of $\mathbf{F}_k$

Notice that $g(\bullet)$ is a decreasing function, and $\gamma_n(\{\mathbf{F}_k\}_{k=1}^K)$ is an increasing function of $\lambda_n(\mathbf{F}_k^H \bar{\mathbf{H}}_k^H \mathbf{K}_{\mathbf{F}_k}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_k)$. Therefore, $g[\boldsymbol{\gamma}(\{\mathbf{F}_k\}_{k=1}^K)]$ is a decreasing matrix-monotone function of $\mathbf{F}_k^H \bar{\mathbf{H}}_k^H \mathbf{K}_{\mathbf{F}_k}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_k$. It can be proved that at the optimal solution, the power constraints hold at the equality, i.e., $\text{Tr}(\mathbf{F}_k \mathbf{F}_k^H) = P_k$, meaning that the relays transmit at the maximum power.

Defining an auxiliary scalar $\eta_{f_k} = \alpha_k \text{Tr}(\mathbf{F}_k \mathbf{F}_k^H \boldsymbol{\Psi}_k) + \sigma_{n_k}^2$ with $\alpha_k = \text{Tr}(\boldsymbol{\Sigma}_k)/M_k$, $\text{Tr}(\mathbf{F}_k \mathbf{F}_k^H) = P_k$ is equivalent to $\text{Tr}[\mathbf{F}_k \mathbf{F}_k^H(\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2 \mathbf{I}_{N_k})]/\eta_{f_k} = P_k$ [11]. Thus the robust transceiver design problem (31) is equivalent to

$$\min_{\mathbf{F}_k} \quad g[\boldsymbol{\gamma}(\{\mathbf{F}_k\}_{k=1}^K)]$$

$$\text{s.t.} \quad \gamma_n(\{\mathbf{F}_k\}_{k=1}^K) = \prod_{k=1}^K \frac{\lambda_n(\mathbf{F}_k^H \bar{\mathbf{H}}_k^H \mathbf{K}_{\mathbf{F}_k}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_k)}{1 + \lambda_n(\mathbf{F}_k^H \bar{\mathbf{H}}_k^H \mathbf{K}_{\mathbf{F}_k}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_k)}$$

$$\text{Tr}[\mathbf{F}_k \mathbf{F}_k^H(\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2 \mathbf{I}_{N_k})]/\eta_{f_k} = P_k. \tag{32}$$

It can be proved that when $\boldsymbol{\Psi}_k \propto \mathbf{I}$ or $\boldsymbol{\Sigma}_k \propto \mathbf{I}$, the optimal solutions of the optimization problem (32) have the following structure [11]

$$\mathbf{F}_{k,\text{opt}} = \sqrt{\xi_k(\boldsymbol{\Lambda}_{\mathcal{F}_k})}(\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2 \mathbf{I})^{-1/2} \mathbf{V}_{\mathcal{H}_k,N} \boldsymbol{\Lambda}_{\mathcal{F}_k}$$
$$\times \mathbf{U}_{\text{Arb}_k,N}^H. \tag{33}$$

In (33), $\xi_k(\boldsymbol{\Lambda}_{\mathcal{F}_k})$ equals to

$$\xi_k(\boldsymbol{\Lambda}_{\mathcal{F}_k}) = \sigma_{n_k}^2 / \{1 - \alpha_k \text{Tr}[\mathbf{V}_{\mathcal{H}_k,N}^H(\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2 \mathbf{I})^{-\frac{1}{2}}$$
$$\times \boldsymbol{\Psi}_k(\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2 \mathbf{I})^{-\frac{1}{2}} \mathbf{V}_{\mathcal{H}_k,N} \boldsymbol{\Lambda}_{\mathcal{F}_k}^2]\} = \eta_{f_k}, \tag{34}$$

where $\boldsymbol{\Lambda}_{\mathcal{F}_k}$ is a $N \times N$ unknown diagonal matrix, and $\mathbf{V}_{\mathcal{H}_k,N}$ and $\mathbf{U}_{\text{Arb}_k,N}$ are the matrices consisting of the first N columns of $\mathbf{V}_{\mathcal{H}_k}$ and $\mathbf{U}_{\text{Arb}_k}$, respectively. The unitary matrix $\mathbf{U}_{\text{Arb}_k}$ is an arbitrary $M_{k-1} \times M_{k-1}$ unitary matrix, and the unitary matrix $\mathbf{V}_{\mathcal{H}_k}$ is defined based on the following singular value decomposition

$$(\mathbf{K}_{\mathbf{F}_k}/\eta_{f_k})^{-1/2} \bar{\mathbf{H}}_k(\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2 \mathbf{I})^{-1/2}$$
$$= \mathbf{U}_{\mathcal{H}_k} \boldsymbol{\Lambda}_{\mathcal{H}_k} \mathbf{V}_{\mathcal{H}_k}^H \tag{35}$$

where the diagonal elements of $\boldsymbol{\Lambda}_{\mathcal{H}_k}$ are arranged in decreasing order.

It is obvious that in (33), the only unknown variable is $\boldsymbol{\Lambda}_{\mathcal{F}_k}$. It can be solved via using iterative water-filling algorithm with guaranteed convergence [11].
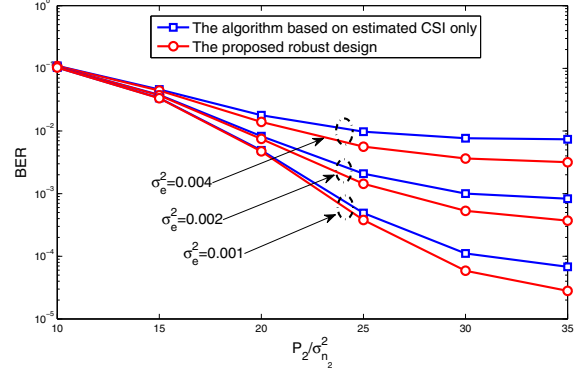

Fig. 1. BERs of proposed robust design with M-Schur-convex objective functions and the algorithm based on estimated CSI only when $\rho_t = 0$, $\rho_r = 0.03$, and $P_1/\sigma_{n_1}^2 = 30$dB.

Remark: The mathematic framework given in this paper can be directly extended to robust transceiver design with decision feedback equalizer (DFE) for multi-hop AF MIMO relaying systems. Furthermore, the derivation logic in this paper can be applicable to robust linear transceiver design for multi-hop AF MIMO relaying systems with Schur-convex/concave (additive majorization) objective functions [13]. Finally, based on **Property 1** the robust transceiver design discussed in this paper is actually a unified framework for transceiver design for multi-hop AF MIMO relaying systems, which is comprehensible and powerful.

VI. SIMULATION RESULTS AND DISCUSSIONS

In this section, the performance of the proposed algorithms is assessed by simulations. In the following, we consider an AF MIMO relaying system where the source, relays and destination are all equipped with four antennas, i.e., $N_k = M_k = 4$. The estimation error correlation matrices are chosen as the popular exponential model $[\boldsymbol{\Psi}_k] = \sigma_e^2 \rho_t^{|i-j|}$ and $[\boldsymbol{\Sigma}_k] = \rho_r^{|i-j|}$ [8] where ρ_t and ρ_r are the correlation coefficients, and σ_e^2 denotes the estimation error variance. The estimated channels $\bar{\mathbf{H}}_k$'s are randomly generated based on the following complex Gaussian distributions $\bar{\mathbf{H}}_k \sim \mathcal{CN}_{M_k,N_k}(\mathbf{0}_{M_k,N_k}, (1 - \sigma_e^2)/\sigma_e^2 \boldsymbol{\Sigma}_k \otimes \boldsymbol{\Psi}_k^T)$ [8], 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 $P_k/\sigma_{n_k}^2$. At the source node, four independent data streams are transmitted and in each data stream, $N_{\text{Data}} = 10^6$ independent 16-QAM symbols are transmitted. Moreover, Gray code is also used to improve system performance further. Each point in the following figures is an average of 10000 trials.

Fig. 1 shows BERs of the proposed robust nonlinear design with the M-Schur convex objective and the corresponding algorithm based on estimated CSI only (which takes the channel estimates as true channels) with $\rho_t = 0$, $\rho_t = 0.03$, $P_1/\sigma_{n_1}^2 = 30$dB, and $P_2/\sigma_{n_2}^2$ being varied from 10dB to 35dB. It can be seen that smaller estimation errors lead to better performance for both algorithms, but the performance

of the proposed algorithm is always better than that based on the estimated CSI only.

VII. Conclusions

Joint Bayesian robust transceiver design for multi-hop AF MIMO relaying systems was investigated, where channel estimation errors exist in CSI in all hops. At the source node, a nonlinear THP was used, and was jointly optimized with linear forwarding matrices at all relays and linear equalizer at the destination. A general transceiver optimization problem was formulated with objective function being either M-Scur-convex or M-Scur-concave. Exploiting the elegant properties of multiplicative majorization theory and matrix-monotone functions, the optimal structure of the transceivers was first derived. Then, the original optimization problem was greatly simplified and an iterative water-filling solution was proposed to solve for the remaining unknown variables. Simulation results showed that the proposed robust design has much better performance than the non-robust design.

Appendix A
Proof of Property 1

First notice two facts in matrix theory: (a) for two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ with compatible dimension $\lambda_i(\boldsymbol{AB}) = \lambda_i(\boldsymbol{BA})$ [12, 9.A.1.a]; (b) for two positive semi-definite matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, $\prod_{i=1}^n \lambda_i(\boldsymbol{AB}) \leq \prod_{i=1}^n \lambda_i(\boldsymbol{A})\lambda_i(\boldsymbol{B})$ [12, 9.H.1.a], where the equality holds when $\boldsymbol{A}$ and $\boldsymbol{B}$ has the same unitary matrix in eigendecomposition [12]. With these two facts, we have

$$\prod_{i=1}^n \lambda_i(\mathbf{M}_1^{\mathrm{H}}\mathbf{Q}_1^{\mathrm{H}}\cdots\mathbf{M}_K^{\mathrm{H}}\mathbf{Q}_K^{\mathrm{H}}\mathbf{Q}_K\mathbf{M}_K\cdots\mathbf{Q}_1\mathbf{M}_1)$$

$$=\prod_{i=1}^n \lambda_i(\mathbf{M}_2^{\mathrm{H}}\mathbf{Q}_2^{\mathrm{H}}\cdots\mathbf{M}_K^{\mathrm{H}}\mathbf{Q}_K^{\mathrm{H}}\mathbf{Q}_K\mathbf{M}_K\cdots\mathbf{Q}_2\mathbf{M}_2\mathbf{Q}_1\mathbf{M}_1\mathbf{M}_1^{\mathrm{H}}\mathbf{Q}_1^{\mathrm{H}})$$

$$\leq\prod_{i=1}^n \lambda_i(\mathbf{M}_2^{\mathrm{H}}\mathbf{Q}_2^{\mathrm{H}}\cdots\mathbf{M}_K^{\mathrm{H}}\mathbf{Q}_K^{\mathrm{H}}\mathbf{Q}_K\mathbf{M}_K\cdots\mathbf{Q}_2\mathbf{M}_2)\lambda_i(\mathbf{M}_1\mathbf{M}_1^{\mathrm{H}})$$

$$n = 1,\cdots,N, \tag{36}$$

where the first equality is due to fact (a) and the second inequality is based on fact (b). Repeating the above two processes and based on the fact that $\lambda_i(\mathbf{M}_k\mathbf{M}_k^{\mathrm{H}}) = \lambda_i(\mathbf{M}_k^{\mathrm{H}}\mathbf{M}_k)$ we can obtain the following inequality

$$\prod_{i=1}^n \lambda_i(\boldsymbol{\Theta})$$

$$\leq \prod_{i=1}^n \underbrace{\lambda_i(\mathbf{M}_K^{\mathrm{H}}\mathbf{M}_K)\lambda_i(\mathbf{M}_{K-1}^{\mathrm{H}}\mathbf{M}_{K-1})\cdots\lambda_i(\mathbf{M}_1^{\mathrm{H}}\mathbf{M}_1)}_{\triangleq\gamma_i(\{\mathbf{F}_k\}_{k=1}^K)},$$

$$\tag{37}$$

where the equality holds when $\mathbf{Q}_k$'s satisfy

$$\mathbf{Q}_k = \mathbf{V}_{\mathbf{M}_{k+1}}\mathbf{U}_{\mathbf{M}_k}^{\mathrm{H}}, \quad k = 1,\cdots,K \tag{38}$$

where $\mathbf{U}_{\mathbf{M}_k}$ and $\mathbf{V}_{\mathbf{M}_k}$ are defined based on the following singular value decomposition $\mathbf{M}_k = \mathbf{U}_{\mathbf{M}_k}\boldsymbol{\Lambda}_{\mathbf{M}_k}\mathbf{V}_{\mathbf{M}_k}^{\mathrm{H}}$ with the diagonal elements of $\boldsymbol{\Lambda}_{\mathbf{M}_k}$ arranged in decreasing order. Furthermore, based on the definition of $\mathbf{M}_k$ in (18), $\gamma_i(\{\mathbf{F}_k\}_{k=1}^K)$ in (37) equals to

$$\gamma_i(\{\mathbf{F}_k\}_{k=1}^K) = \prod_{k=1}^K \frac{\lambda_i(\mathbf{F}_k^{\mathrm{H}}\bar{\mathbf{H}}_k^{\mathrm{H}}\mathbf{K}_{\mathbf{F}_k}^{-1}\bar{\mathbf{H}}_k\mathbf{F}_k)}{1 + \lambda_i(\mathbf{F}_k^{\mathrm{H}}\bar{\mathbf{H}}_k^{\mathrm{H}}\mathbf{K}_{\mathbf{F}_k}^{-1}\bar{\mathbf{H}}_k\mathbf{F}_k)}. \tag{39}$$

References

[1] O. Munoz-Medina, J. Vidal, and A. Agustin, "Linear transceiver design in nonregenerative relays with channel state information," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2593–2604, June 2007.

[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] 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, July 2008.

[4] 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.

[5] Y. Jiang, J. Li, and W. W. Hager, "Uniform channel decomposition for MIMO communications," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4283–4294, Nov. 2005.

[6] A. A. D'Amico, "Tomlinson-Harashima precoding in MIMO systems: A unified approach to transceiver optimization based on multiplicative Schur-convexity," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3662–3677, Aug. 2008.

[7] Y. Rong and M. R. A. Khandaken, "Uplink-downlink duality of multi-hop MIMO relay channel," *IEEE Trans. Wireless Commun.*, vol. 10, no. 6, pp. 1923–1931, June 2011.

[8] 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.

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

[10] R. F. H. Fischer, *Precoding and Signal Shaping for Digital Transmission*. New York: Wiley-IEEE, July 2002.

[11] C. Xing, M. Xia, F. Gao, and Y.-C. Wu, "Robust transceiver with Tomlinson-Harashima precoding for amplify-and-forward MIMO relaying systems," *Submitted to JSAC*, available at: http://arxiv.org/PS_cache/arxiv/pdf/1109/1109.0847v2.pdf.

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

[13] C. Xing, Z. Fei, S. Ma, J. Kuang, and Y.-C. Wu, "Robust linear transceiver design for non-regenerative AF MIMO relaying systems," WCSP 2011, Nanjing, China, Nov. 2011.