Title	Robust Transceiver with Tomlinson-Harashima Precoding for Amplify-and-Forward MIMO Relaying Systems
Author(s)	Xing, C; Xia, M; Gao, F; Wu, YC
Citation	IEEE Journal on Selected Areas in Communications, 2012, v. 30 n. 8, p. 1370-1382
Issued Date	2012
URL	http://hdl.handle.net/10722/189016
Rights	IEEE Journal on Selected Areas in Communications. Copyright © IEEE.

Robust Transceiver with Tomlinson-Harashima Precoding for Amplify-and-Forward MIMO Relaying Systems

Chengwen Xing, Minghua Xia, Feifei Gao, and Yik-Chung Wu

Abstract—In this paper, robust transceiver design with Tomlinson-Harashima precoding (THP) for multi-hop amplify-and-forward (AF) multiple-input multiple-output (MIMO) relaying systems is investigated. At source node, THP is adopted to mitigate the spatial intersymbol interference. However, due to its nonlinear nature, THP is very sensitive to channel estimation errors. In order to reduce the effects of channel estimation errors, a joint Bayesian robust design of THP at source, linear forwarding matrices at relays and linear equalizer at destination is proposed. With novel applications of 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 transceiver design problem reduces to a much simpler one with only scalar variables which can be efficiently solved. Finally, the performance advantage of the proposed robust design over non-robust design is demonstrated by simulation results.

Index Terms—Amplify-and-forward (AF), multiple-input multiple-output (MIMO), Tomlinson-Harashima precoding, robust design, majorization theory.

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. AF MIMO relaying systems have a broad range of potential applications including resource exploration, vehicle communications, military ad hoc networks, satellite communications, etc [1]. This system has also been considered to be adopted in the emerging wireless systems, such as LTE-Advanced and WINNER project.

Linear transceiver design for dual-hop AF MIMO relaying systems has been extensively studied in [2]–[12]. In particular, joint design of relay forwarding matrix and destination equalizer minimizing mean-square-error (MSE) of data streams

C. Xing is with the School of Information and Electronics, Beijing Institute of Technology, Beijing, China (e-mail: chengwenxing@ieee.org).

M. Xia is with the Division of Physical Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia (e-mail: minghua.xia@ieee.org).

F. Gao is with Tsinghua National Laboratory for Information Science and Technology, Beijing, China (e-mail: feifeigao@ieee.org).

Y.-C. Wu is with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong (e-mail: ycwu@eee.hku.hk).

Digital Object Identifier 10.1109/JSAC.2012.120907.

is discussed in [4]. Joint design of source precoder, relay forwarding matrix and destination equalizer minimizing MSE is investigated in [5], [6], [9]. The capacity maximization transceiver design has also been reported in [2], [3], [9]. On the other hand, linear transceiver design for multi-hop AF MIMO relaying systems with prefect channel state information (CSI) is discussed in [12]. Furthermore, robust design, which takes channel estimation errors into account, is recently investigated in [7], [8], [10], [11], where the channel estimation uncertainty is considered as nuisance parameters and removed in Bayesian sense.

In general, there are two goals in transceiver designs: transmitting as much information as possible and recovering the signal at receiver as accurately as possible. The latter one is the starting point of this paper. For multiple-antenna systems with fixed bit rates, it is well-known that nonlinear transceivers usually have performance advantage in terms of bit error rate (BER) than their linear counterparts [13]–[15]. Recently, nonlinear transceiver design for AF MIMO relaying systems assuming perfect CSI, was introduced in [16]. There are two kinds of nonlinear transceiver design: decision-feedback equalization (DFE) based design and Tomlinson-Harashima precoding (THP) based design. In fact, there exists a duality between these two designs [16], [17]. 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 [18]. Therefore, robust nonlinear transceiver design is a promising way to mitigate such problem. This is the motivation of the current work.

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 in this case 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 multiplicative Schur-convex/concave. With novel applications of results in multiplicative Schur-convexity and matrix-monotone functions, the optimal diagonal structure of
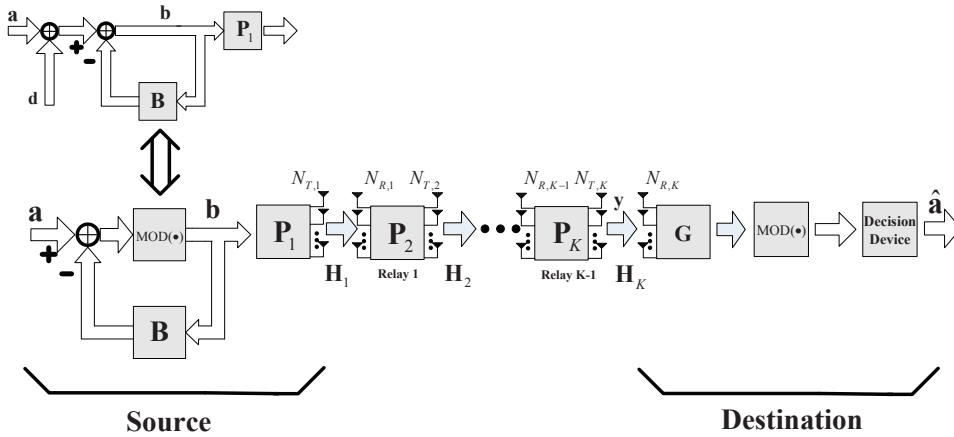
Fig. 1.   Multi-hop AF MIMO relaying system with Tomlinson-Harashima Precoding at the source.

the transceiver is derived. With the obtained optimal structures, the transceiver design is then significantly simplified and then iterative water-filling alike solutions are adopted to solve for the remaining unknown variables. It is found that if the objective function is multiplicatively Schur-concave, the proposed nonlinear transceiver design reduces to linear transceiver design. 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. Notice that while delay is a critical consideration for relaying communication, in this paper, we assume that the network size is limited and the effects of time delay in transmission are not considered.

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 $\mathbf{I}_N$ denotes an $N \times N$ 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 $[\mathbf{Z}]_{n,m}$ and $\boldsymbol{\lambda}(\mathbf{Z})$ represents the vector consisting of the eigenvalues of $\mathbf{Z}$.

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, as shown in Fig. 1. The source is equipped with $N_{T,1}$ transmit antennas. The kth relay has $N_{R,k}$ receive antennas and $N_{T,k+1}$ transmit antennas. The destination is equipped with $N_{R,K}$ receive antennas. At the source, at each time slot, there is a $N \times 1$ vector $\mathbf{a} = [a_1, a_2, \cdots, a_N]^{\mathrm{T}}$ 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 $\mathcal{A} = \{\pm 1, \pm 3, \cdots, \pm(\sqrt{\mathcal{M}} - 1)\}$ [1].

As shown by Fig. 1, at the transmitter, the data vector $\mathbf{a}$ is fed into the a precoding unit which consists of a

[1]In this paper, only square QAM is considered.

$N \times N$ feedback matrix $\mathbf{B}$ and a nonlinear modulo operator $\mathrm{MOD}_{\mathcal{M}}(\bullet)$. The square matrix $\mathbf{B}$ is a strictly lower triangular matrix which allows data precoding in a recursive fashion and the $\mathrm{MOD}_{\mathcal{M}}(\bullet)$ is defined as

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

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 $[-\sqrt{\mathcal{M}}, \sqrt{\mathcal{M}}] \times [-\sqrt{\mathcal{M}}, \sqrt{\mathcal{M}}]$. In the equation, $\mathrm{Re}(x)$ and $\mathrm{Im}(x)$ denote the real and imaginary parts of x, respectively.

Generally speaking, nonlinear operation is more complicated to be analyzed than linear operation. To simplify the following analysis, as shown by Fig. 1, 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 $d_k = 2\sqrt{\mathcal{M}}I_k$ 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 $[-\sqrt{\mathcal{M}}, \sqrt{\mathcal{M}}] \times [-\sqrt{\mathcal{M}}, \sqrt{\mathcal{M}}]$. 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 $\mathbf{b} \triangleq [b_1, \cdots, b_N]^{\mathrm{T}}$, $\mathbf{d} \triangleq [d_1, \cdots, d_N]^{\mathrm{T}}$, 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 $\mathrm{Tr}(\mathbf{P}_1\mathbf{R}_b\mathbf{P}_1^{\mathrm{H}}) \leq P_1$ where P_1 is the maximum transmit power at the source. When the elements of $\mathbf{a}$ are independent and identically distributed (i.i.d.) over the constellation and the dimension of modulation constellation $\mathcal{M}$ is large, $\mathbf{b}$ can be considered as i.i.d. [19], i.e.,

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

The received signal $\mathbf{x}_1$ at the first relay is formulated as

$$\mathbf{x}_1 = \mathbf{H}_1\mathbf{P}_1\mathbf{b} + \mathbf{n}_1 \quad (5)$$

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 $\mathbf{R}_{\mathbf{n}_1} = \sigma_{n_1}^2 \mathbf{I}_{N_{R,1}}$.

At the first relay, the received signal $\mathbf{x}_1$ is multiplied by a forwarding matrix $\mathbf{P}_2$ and then the resultant signal is transmitted to the second relay. The received signal at the second relay can be written as

$$\mathbf{x}_2 = \mathbf{H}_2\mathbf{P}_2\mathbf{H}_1\mathbf{P}_1\mathbf{b} + \mathbf{H}_2\mathbf{P}_2\mathbf{n}_1 + \mathbf{n}_2 \quad (6)$$

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

$$\mathbf{x}_k = \mathbf{H}_k\mathbf{P}_k\mathbf{x}_{k-1} + \mathbf{n}_k \quad (7)$$

with $\mathbf{H}_k$ and $\mathbf{n}_k$ are the channel and additive noise at the kth hop, respectively. In this paper, we considered slow fading channels with $\mathbf{H}_k$ being fixed in each transmission.

The covariance matrix of $\mathbf{n}_k$ is denoted as $\mathbf{R}_{\mathbf{n}_k} = \sigma_{n_k}^2 \mathbf{I}_{N_{R,k}}$. Finally, for a K-hop AF MIMO relaying system, the received signal at the destination is

$$\mathbf{y} = \left[\prod_{k=1}^{K}(\mathbf{H}_k\mathbf{P}_k)\right]\mathbf{b} + \sum_{k=1}^{K-1}\left\{\left[\prod_{l=k+1}^{K}(\mathbf{H}_l\mathbf{P}_l)\right]\mathbf{n}_k\right\} + \mathbf{n}_K, \quad (8)$$

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 $N_{T,k}$ and $N_{R,k}$ are greater than or equal to N [4].

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

$$\mathbf{H}_k = \bar{\mathbf{H}}_k + \Delta\mathbf{H}_k, \quad (9)$$

where $\bar{\mathbf{H}}_k$ is the estimated channels, and $\Delta\mathbf{H}_k$ is the corresponding channel estimation errors[2] whose elements are zero mean Gaussian random variables. Furthermore, the $N_{R,k} \times N_{T,k}$ matrix $\Delta\mathbf{H}_k$ can be decomposed using the widely used Kronecker model [7], [8], [20] as $\Delta\mathbf{H}_k = \boldsymbol{\Sigma}_k^{1/2}\mathbf{H}_{W,k}\boldsymbol{\Psi}_k^{1/2}$, where the elements of the $N_{R,k} \times N_{T,k}$ matrix $\mathbf{H}_{W,k}$ are i.i.d. Gaussian random variables with zero mean and unit variance. The specific formulas of $\boldsymbol{\Sigma}_k$ and $\boldsymbol{\Psi}_k$ are determined by the training sequences and channel estimators [7], [8], [11], [21].

B. Problem Formulation

As shown by Fig. 1, at the destination, a linear equalizer $\mathbf{G}$ is adopted and is followed by a modulo operator. As the real and imaginary parts of $\mathbf{d}$ are both integer multiples of $2\sqrt{\mathcal{M}}$, the effect of $\mathbf{d}$ will be perfectly removed by modulo operator at the destination. As a result, estimating $\mathbf{s}$ is equivalent to estimating $\mathbf{a}$ [14], [15]. 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 $\mathbb{E}\{(\mathbf{Gy}-\mathbf{s})(\mathbf{Gy}-\mathbf{s})^{\text{H}}\}$ [15], [19], 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

$$\begin{aligned}
&\boldsymbol{\Phi}(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^K, \mathbf{C}) \\
&= \mathbb{E}\{(\mathbf{Gy} - \mathbf{Cb})(\mathbf{Gy} - \mathbf{Cb})^{\text{H}}\} \\
&= \mathbf{G}[\bar{\mathbf{H}}_K\mathbf{P}_K\mathbf{R}_{\mathbf{x}_{K-1}}\mathbf{P}_K^{\text{H}}\bar{\mathbf{H}}_K^{\text{H}} + \text{Tr}(\mathbf{P}_K\mathbf{R}_{\mathbf{x}_{K-1}}\mathbf{P}_K^{\text{H}}\boldsymbol{\Psi}_K)\boldsymbol{\Sigma}_K \\
&\quad + \mathbf{R}_{\mathbf{n}_K}]\mathbf{G}^{\text{H}} - \sigma_b^2\mathbf{G}\prod_{k=1}^{K}\left(\bar{\mathbf{H}}_k\mathbf{P}_k\right)\mathbf{C}^{\text{H}} \\
&\quad - \sigma_b^2\left[\mathbf{G}\prod_{k=1}^{K}\left(\bar{\mathbf{H}}_k\mathbf{P}_k\right)\mathbf{C}^{\text{H}}\right]^{\text{H}} + \sigma_b^2\mathbf{C}\mathbf{C}^{\text{H}} \quad (10)
\end{aligned}$$

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

$$\begin{aligned}
\mathbf{R}_{\mathbf{x}_k} &\triangleq \mathbb{E}\{\mathbf{x}_k\mathbf{x}_k^{\text{H}}\} \\
&= \bar{\mathbf{H}}_k\mathbf{P}_k\mathbf{R}_{\mathbf{x}_{k-1}}\mathbf{P}_k^{\text{H}}\bar{\mathbf{H}}_k^{\text{H}} + \text{Tr}(\mathbf{P}_k\mathbf{R}_{\mathbf{x}_{k-1}}\mathbf{P}_k^{\text{H}}\boldsymbol{\Psi}_k)\boldsymbol{\Sigma}_k + \mathbf{R}_{\mathbf{n}_k}.
\end{aligned} \quad (11)$$

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 $f([\text{MSE}_1, \cdots, \text{MSE}_N]^{\text{T}}) = \sum_{n=1}^{N}\text{MSE}_n$, where $\text{MSE}_n = [\boldsymbol{\Phi}(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^K, \mathbf{C})]_{n,n}$. For product MSE minimization, the objective function is $f([\text{MSE}_1, \cdots, \text{MSE}_N]^{\text{T}}) = \prod_{n=1}^{N}\text{MSE}_n$. Furthermore, worst-case MSE minimization corresponds to minimizing the objective function given as $f([\text{MSE}_1, \cdots, \text{MSE}_N]^{\text{T}}) = \max_{n=1,2,\cdots,N}\{\text{MSE}_n\}$ [9], [13], [15], [23]. On the other hand, weighted geometric mean MSE minimization corresponds to minimizing the following objective function $f([\text{MSE}_1, \cdots, \text{MSE}_N]^{\text{T}}) = \prod_{n=1}^{N}\text{MSE}_n^{w_n}$ with $w_1 \geq w_2 \cdots \geq w_N \geq 0$. Therefore, a unified transceiver design optimization problem can be formulated as

$$\begin{aligned}
\min_{\mathbf{G}, \mathbf{P}_k, \mathbf{C}} \quad & f([\text{MSE}_1, \cdots, \text{MSE}_N]^{\text{T}}) \\
\text{s.t.} \quad & \text{MSE}_n = [\boldsymbol{\Phi}(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^K, \mathbf{C})]_{n,n} \\
& \text{Tr}(\mathbf{P}_k\mathbf{R}_{\mathbf{x}_{k-1}}\mathbf{P}_k^{\text{H}}) \leq P_k, \quad k = 1, \cdots, K \quad (12)
\end{aligned}$$

where the matrix $\mathbf{C}$ is a lower triangular matrix with unit diagonal elements and P_k is the maximum transmit power at the kth node.

In general, the objective function $f(\bullet)$ possesses two important properties:
(1) $f(\bullet)$ is an increasing real-valued vector function $\mathbb{R}^N \to \mathbb{R}$, i.e., for two vectors $\mathbf{u} = [u_1, u_2, \cdots, u_N]^{\text{T}}$ and $\mathbf{v} = [v_1, v_2, \cdots, v_N]^{\text{T}}$, when $u_n \geq v_n$, we have $f(\mathbf{u}) \geq f(\mathbf{v})$. This property is natural in transceiver design. This is because for two designs resulting in $[\text{MSE}_1, \cdots, \text{MSE}_N]^{\text{T}}$ and $[\widehat{\text{MSE}_1}, \cdots, \widehat{\text{MSE}_N}]^{\text{T}}$, suppose $\text{MSE}_n < \widehat{\text{MSE}_n}$ for all n, we will prefer the former design. This fact is reflected in $f(\bullet)$ being an increasing function.
(2) $f(\bullet)$ is multiplicatively Schur-convex or concave, with definitions given below.

Definition 1: For any $\mathbf{z} \in \mathbb{R}^n$, let $z_{[k]}$ denotes the kth largest elements of $\mathbf{z}$ and $z_{(k)}$ denotes the kth smallest elements of $\mathbf{z}$,

---

[2]In this paper, only channel estimation errors are taken into account.

i.e., $z_{[1]} \geq \cdots \geq z_{[N]}$ and $z_{(1)} \leq \cdots \leq z_{(N)}$. For two vectors $\mathbf{v}, \mathbf{u}$ whose elements are **nonnegative**, $\mathbf{v} \prec_{\times} \mathbf{u}$ is defined as

$$\prod_{i=1}^{k} v_{[i]} \leq \prod_{i=1}^{k} u_{[i]}, \; k = 1, \cdots, N-1 \text{ and } \prod_{i=1}^{N} v_{[i]} = \prod_{i=1}^{N} u_{[i]}. \quad (13)$$

Definition 2: A function $\phi(\bullet)$ is multiplicatively Schur-convex if and only if $\mathbf{v} \prec_{\times} \mathbf{u}$ implies $\phi(\mathbf{v}) \leq \phi(\mathbf{u})$. Notice that $\phi(\bullet)$ is multiplicatively Schur-convex if and only if $-\phi(\bullet)$ is multiplicatively Schur-concave.

Notice that **Definition 2** cannot be directly used to prove whether a function is multiplicatively Schur-convex or Schur-concave. In practice, we need the following **Lemma 1**.

Lemma 1: Let $\phi(\bullet)$ be a continuous real-valued function defined on $\mathcal{D} = \{\mathbf{z} : z_1 \geq \cdots \geq z_N \geq 0\}$. Then $\phi(\bullet)$ is multiplicatively Schur-convex if and only if for all $\mathbf{z} \in \mathcal{D}$,

$$\phi(z_1, \cdots, z_{k-1}, z_k/e, z_{k+1} \times e, z_{k+2}, \cdots, z_N)$$

is decreasing in e over the following regions

$$1 \leq e \text{ and } z_k/e \geq z_{k+1} \times e \text{ for } k = 1, \cdots, N-1. \quad (14)$$

Proof: See Appendix A. ∎

With **Lemma 1** and straightforward computation, it can be proved that the four objective functions mentioned above are multiplicatively Schur-convex or concave. In the following, for notational convenience, multiplicatively Schur-convex/concave is referred to as M-Schur-convex/concave.

Remark 1: Notice that in [14], [15], there is another way to prove whether a function is M-Schur-convex/concave. However, the method in [14], [15] requires all input variables $z_1, z_2, ..., z_N > 0$. In contrast, **Lemma 1** provides a stronger result and allows elements of $\mathbf{z}$ being zero.

Remark 2: The differences between our work and [14], [15] are twofold. (a) The system considered in [14], [15] is a point-to-point MIMO system, while our work focuses on a multi-hop AF MIMO relaying system. (b) In the above two works, the involved CSI is perfectly known. In this paper, we consider a robust transceiver design under Gaussian distributed channel estimation errors. Generally speaking, the problem tackled in this paper is more complicated and more challenging, because of more variables, more constraints, a more complicated objective function.

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 (10) with respect to $\mathbf{G}^*$ (the conjugate of $\mathbf{G}$) to be zero, and we have

$$\mathbf{G}_{\text{LMMSE}} = \sigma_b^2 \left[\prod_{k=1}^{K} (\bar{\mathbf{H}}_k \mathbf{P}_k) \mathbf{C}^{\text{H}} \right]^{\text{H}} [\bar{\mathbf{H}}_K \mathbf{P}_K \mathbf{R}_{\mathbf{x}_{K-1}} \mathbf{P}_K^{\text{H}} \bar{\mathbf{H}}_K^{\text{H}}$$
$$+ \text{Tr}(\mathbf{P}_K \mathbf{R}_{\mathbf{x}_{K-1}} \mathbf{P}_K^{\text{H}} \mathbf{\Psi}_K) \mathbf{\Sigma}_K + \mathbf{R}_{\mathbf{n}_K}]^{-1}. \quad (15)$$

In terms of MSE, LMMSE estimator is a dominated estimator in linear estimators [23], i.e.,

$$\mathbf{\Phi}(\mathbf{G}_{\text{LMMSE}}, \{\mathbf{P}_k\}_{k=1}^{K}, \mathbf{C}) \preceq \mathbf{\Phi}(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^{K}, \mathbf{C}) \quad (16)$$

which implies

$$[\mathbf{\Phi}(\mathbf{G}_{\text{LMMSE}}, \{\mathbf{P}_k\}_{k=1}^{K}, \mathbf{C})]_{n,n} \leq [\mathbf{\Phi}(\mathbf{G}, \{\mathbf{P}_k\}_{k=1}^{K}, \mathbf{C})]_{n,n}.$$

As $f(\bullet)$ is an increasing function, and there is no constraint on $\mathbf{G}$ in (12), the optimal linear equalizer is LMMSE equalizer, i.e., $\mathbf{G}_{\text{opt}} = \mathbf{G}_{\text{LMMSE}}$.

Substituting the optimal equalizer (15) into the MSE formulation (10), the MSE matrix is rewritten as

$$\mathbf{\Phi}_{\text{MSE}}(\{\mathbf{P}_k\}_{k=1}^{K}, \mathbf{C})$$
$$= \sigma_b^2 \mathbf{C} \left(\mathbf{I}_N - \sigma_b^2 \left[\prod_{k=1}^{K} (\bar{\mathbf{H}}_k \mathbf{P}_k) \mathbf{C}^{\text{H}} \right]^{\text{H}} [\bar{\mathbf{H}}_K \mathbf{P}_K \mathbf{R}_{\mathbf{x}_{K-1}} \right.$$
$$\times \mathbf{P}_K^{\text{H}} \bar{\mathbf{H}}_K^{\text{H}} + \text{Tr}(\mathbf{P}_K \mathbf{R}_{\mathbf{x}_{K-1}} \mathbf{P}_K^{\text{H}} \mathbf{\Psi}_K) \mathbf{\Sigma}_K + \mathbf{R}_{\mathbf{n}_K}]^{-1}$$
$$\left. \times \left[\prod_{k=1}^{K} (\bar{\mathbf{H}}_k \mathbf{P}_k) \mathbf{C}^{\text{H}} \right] \right) \mathbf{C}^{\text{H}} \quad (17)$$

based on which the optimization problem (12) becomes

$$\min_{\mathbf{P}_k, \mathbf{C}} \quad f([\text{MSE}_1, \cdots, \text{MSE}_N]^{\text{T}})$$
$$\text{s.t.} \quad \text{MSE}_n = [\mathbf{\Phi}_{\text{MSE}}(\{\mathbf{P}_k\}_{k=1}^{K}, \mathbf{C})]_{n,n}$$
$$\text{Tr}(\mathbf{P}_k \mathbf{R}_{\mathbf{x}_{k-1}} \mathbf{P}_k^{\text{H}}) \leq P_k. \quad (18)$$

From the definition of $\mathbf{R}_{\mathbf{x}_k}$ in (11), 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 in (18) are coupled with each other. In order to simplify the analysis, we define the following new variables

$$\mathbf{F}_1 = \mathbf{P}_1 \mathbf{R}_{\mathbf{b}}^{1/2} \mathbf{Q}_0^{\text{H}} \quad (19)$$

and $\quad \mathbf{F}_k = \mathbf{P}_k \mathbf{K}_{\mathbf{F}_{k-1}}^{1/2} (\mathbf{K}_{\mathbf{F}_{k-1}}^{-1/2} \bar{\mathbf{H}}_{k-1} \mathbf{F}_{k-1} \mathbf{F}_{k-1}^{\text{H}} \bar{\mathbf{H}}_{k-1}^{\text{H}} \mathbf{K}_{\mathbf{F}_{k-1}}^{-1/2}$

$$+ \mathbf{I}_{N_{R,k-1}})^{1/2} \mathbf{Q}_{k-1}^{\text{H}} \quad (20)$$

where $\mathbf{K}_{\mathbf{F}_k}$ is defined as[3]

$$\mathbf{K}_{\mathbf{F}_k} \triangleq \text{Tr}(\mathbf{F}_k \mathbf{F}_k^{\text{H}} \mathbf{\Psi}_k) \mathbf{\Sigma}_k + \sigma_{n_k}^2 \mathbf{I}_{N_{R,k}}, \quad (21)$$

and the matrix $\mathbf{Q}_k$ is an additional unknown unitary matrix. Based on the definition of $\mathbf{F}_k$ in (19) and (20), it is easy to show that $\mathbf{F}_k \mathbf{F}_k^{\text{H}} = \mathbf{P}_k \mathbf{R}_{\mathbf{x}_{k-1}} \mathbf{P}_k^{\text{H}}$ and thus the power constraints becomes

$$\text{Tr}(\mathbf{P}_k \mathbf{R}_{\mathbf{x}_{k-1}} \mathbf{P}_k^{\text{H}}) = \text{Tr}(\mathbf{F}_k \mathbf{F}_k^{\text{H}}) \leq P_k. \quad (22)$$

Therefore, in terms of the new variables $\mathbf{F}_k$, the power constraints become independent of each other, which facilitates further manipulations.

Meanwhile, using (19) and (20), the MSE matrix is further reformulated as (23) on the top of the next page. Based on (22) and (23), the optimization problem (18) is simplified as

$$\min_{\mathbf{F}_k, \mathbf{Q}_k, \mathbf{C}} \quad f([\text{MSE}_1, \cdots, \text{MSE}_N]^{\text{T}})$$
$$\text{s.t.} \quad \text{MSE}_n = \sigma_b^2 [\mathbf{C}(\mathbf{I}_N - \mathbf{Q}_0^{\text{H}} \mathbf{\Theta} \mathbf{Q}_0) \mathbf{C}^{\text{H}}]_{n,n}$$
$$\mathbf{\Theta} = \mathbf{M}_1^{\text{H}} \mathbf{Q}_1^{\text{H}} \cdots \mathbf{M}_K^{\text{H}} \mathbf{Q}_K^{\text{H}} \mathbf{Q}_K \mathbf{M}_K \cdots \mathbf{Q}_1 \mathbf{M}_1$$
$$\text{Tr}(\mathbf{F}_k \mathbf{F}_k^{\text{H}}) \leq P_k, \quad \mathbf{Q}_k^{\text{H}} \mathbf{Q}_k = \mathbf{I}_{N_{R,k}}. \quad (24)$$

Notice that the largest singular value of $\mathbf{M}_k$ is smaller than one. Therefore, the largest eigenvalue of $\mathbf{\Theta}$ is smaller than

---

[3]Putting the definition of $\mathbf{F}_k$ into (11) and comparing (11) with (21), the matrix $\mathbf{K}_{\mathbf{F}_k}$ can be interpreted as the equivalent noise covariance matrix at the kth hop.

$$\Phi_{\mathrm{MSE}}(\mathbf{Q}_k, \{\mathbf{F}_k\}_{k=1}^K, \mathbf{C}) = \mathbf{C}\left(\mathbf{I}_N - \mathbf{Q}_0^{\mathrm{H}}\left\{\prod_{k=1}^K [\mathbf{Q}_k (\mathbf{K}_{\mathbf{F}_k}^{-1/2}\bar{\mathbf{H}}_k \mathbf{F}_k \mathbf{F}_k^{\mathrm{H}} \bar{\mathbf{H}}_k^{\mathrm{H}} \mathbf{K}_{\mathbf{F}_k}^{-1/2} + \mathbf{I}_{N_{R,k}})^{-1/2} \mathbf{K}_{\mathbf{F}_k}^{-1/2}\bar{\mathbf{H}}_k \mathbf{F}_k]\right\}^{\mathrm{H}}\right.$$

$$\times \left. \left\{\prod_{k=1}^K [\mathbf{Q}_k \underbrace{(\mathbf{K}_{\mathbf{F}_k}^{-1/2}\bar{\mathbf{H}}_k \mathbf{F}_k \mathbf{F}_k^{\mathrm{H}} \bar{\mathbf{H}}_k^{\mathrm{H}} \mathbf{K}_{\mathbf{F}_k}^{-1/2} + \mathbf{I}_{N_{R,k}})^{-1/2} \mathbf{K}_{\mathbf{F}_k}^{-1/2}\bar{\mathbf{H}}_k \mathbf{F}_k}_{\triangleq \mathbf{M}_k}]\right\}\mathbf{Q}_0\right)\mathbf{C}^{\mathrm{H}}\sigma_b^2$$

$$= \sigma_b^2 \mathbf{C}(\mathbf{I}_N - \mathbf{Q}_0^{\mathrm{H}}\underbrace{\mathbf{M}_1^{\mathrm{H}}\mathbf{Q}_1^{\mathrm{H}}\mathbf{M}_2^{\mathrm{H}}\mathbf{Q}_2^{\mathrm{H}}\cdots\mathbf{M}_K^{\mathrm{H}}\mathbf{Q}_K^{\mathrm{H}}\mathbf{Q}_K\mathbf{M}_K\cdots\mathbf{Q}_2\mathbf{M}_2\mathbf{Q}_1\mathbf{M}_1}_{\triangleq\boldsymbol{\Theta}}\mathbf{Q}_0)\mathbf{C}^{\mathrm{H}}. \quad (23)$$

one (see Appendix E) and then $\mathbf{I}_N - \mathbf{Q}_0^{\mathrm{H}}\boldsymbol{\Theta}\mathbf{Q}_0$ is a positive definite matrix. With the Cholesky factorization

$$(\mathbf{I}_N - \mathbf{Q}_0^{\mathrm{H}}\boldsymbol{\Theta}\mathbf{Q}_0)\sigma_b^2 = \mathbf{L}\mathbf{L}^{\mathrm{H}} \quad (25)$$

where $\mathbf{L}$ is a lower triangular matrix, and the definition of MSE_n in the second line of (24), we have

$$\begin{aligned}\mathrm{MSE}_n &= \sigma_b^2[\mathbf{C}(\mathbf{I}_N - \mathbf{Q}_0^{\mathrm{H}}\boldsymbol{\Theta}\mathbf{Q}_0)\mathbf{C}^{\mathrm{H}}]_{n,n}\\ &= ([\mathbf{C}^{\mathrm{H}}]_{:,n})^{\mathrm{H}}\mathbf{L}\mathbf{L}^{\mathrm{H}}[\mathbf{C}^{\mathrm{H}}]_{:,n}\\ &= \sum_{i=1}^{n-1}[\mathbf{L}]_{i,i}^2|[(\mathbf{C}\mathbf{L}\mathbf{D}^{-1})^{\mathrm{H}}]_{i,n}|^2 + [\mathbf{L}]_{n,n}^2\\ &\geq [\mathbf{L}]_{n,n}^2, \end{aligned} \quad (26)$$

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

$$\mathbf{D} = \mathrm{diag}\{[\mathbf{L}_{1,1}, \cdots, \mathbf{L}_{N,N}]^{\mathrm{T}}\}. \quad (27)$$

In order to make the equality in the final line of (26) to hold, we need $\sum_{i=1}^{n-1}[\mathbf{L}]_{i,i}^2|[(\mathbf{C}\mathbf{L}\mathbf{D}^{-1})^{\mathrm{H}}]_{i,n}|^2 = 0$, whose solution is

$$\mathbf{C}_{\mathrm{opt}} = \mathbf{D}\mathbf{L}^{-1}. \quad (28)$$

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

$$\begin{aligned}\min_{\mathbf{F}_k, \mathbf{Q}_k} \quad & f([[\mathbf{L}]_{1,1}^2, \cdots, [\mathbf{L}]_{N,N}^2]^{\mathrm{T}})\\ \mathrm{s.t.} \quad & \sigma_b^2(\mathbf{I}_N - \mathbf{Q}_0^{\mathrm{H}}\boldsymbol{\Theta}\mathbf{Q}_0) = \mathbf{L}\mathbf{L}^{\mathrm{H}}\\ & \boldsymbol{\Theta} = \mathbf{M}_1^{\mathrm{H}}\mathbf{Q}_1^{\mathrm{H}}\cdots\mathbf{M}_K^{\mathrm{H}}\mathbf{Q}_K^{\mathrm{H}}\mathbf{Q}_K\mathbf{M}_K\cdots\mathbf{Q}_1\mathbf{M}_1\\ & \mathrm{Tr}(\mathbf{F}_k\mathbf{F}_k^{\mathrm{H}}) \leq P_k, \quad \mathbf{Q}_k^{\mathrm{H}}\mathbf{Q}_k = \mathbf{I}_{N_{R,k}}. \end{aligned} \quad (29)$$

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

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

Because the objective function of the optimization problem (29) is M-Schur-convex or M-Schur-concave. In the following, we will discuss the two cases separately.

M-Schur-convex:

Taking the determinant on both sides of (25), we have

$$|\sigma_b^2(\mathbf{I}_N - \mathbf{Q}_0^{\mathrm{H}}\boldsymbol{\Theta}\mathbf{Q}_0)| = \prod_{n=1}^N[\mathbf{L}]_{n,n}^2 = \sigma_b^{2N}\prod_{n=1}^N(1 - \lambda_n(\boldsymbol{\Theta})) \quad (30)$$

where $\lambda_n(\boldsymbol{\Theta})$ is the nth largest eigenvalue of $\boldsymbol{\Theta}$. Based on (30), the following multiplicative majorization relationship can be established [24]

$$\sigma_b^2\left[\prod_{n=1}^N(1 - \lambda_n(\boldsymbol{\Theta}))\right]^{\frac{1}{N}} \otimes \mathbf{1}_N \prec_{\times} [[\mathbf{L}]_{1,1}^2, \cdots, [\mathbf{L}]_{N,N}^2]^{\mathrm{T}}, \quad (31)$$

where the symbol $\otimes$ denotes the Kronecker product and $\mathbf{1}_N$ is a $N \times 1$ all-one vector. With **Definition 2** and $f(\bullet)$ being a M-Schur-convex function, (31) leads to

$$f([[\mathbf{L}]_{1,1}^2, \cdots, [\mathbf{L}]_{N,N}^2]^{\mathrm{T}}) \geq \underbrace{f\left(\sigma_b^2\left[\prod_{n=1}^N(1-\lambda_n(\boldsymbol{\Theta}))\right]^{\frac{1}{N}} \otimes \mathbf{1}_N\right)}_{\triangleq g[\boldsymbol{\lambda}(\boldsymbol{\Theta})]}, \quad (32)$$

where $\boldsymbol{\lambda}(\boldsymbol{\Theta}) = [\lambda_1(\boldsymbol{\Theta}), \cdots, \lambda_N(\boldsymbol{\Theta})]^{\mathrm{T}}$. The equality in (32) holds when $\prec_{\times}$ in (31) is replaced by equality, which means that $[\mathbf{L}]_{n,n}^2$ are identical for all n. Notice that from (25), we can write $\mathbf{L}\mathbf{L}^{\mathrm{H}} = \sigma_b^2\mathbf{Q}_0^{\mathrm{H}}(\mathbf{I} - \boldsymbol{\Theta})\mathbf{Q}_0$. Since $\mathbf{I} - \boldsymbol{\Theta}$ is positive definite, there always exists an unitary matrix $\mathbf{Q}_0$ which makes the Cholesky factorization matrix of $\mathbf{Q}_0^{\mathrm{H}}(\mathbf{I} - \boldsymbol{\Theta})\mathbf{Q}_0$ have identical diagonal elements [15]. An explicit algorithm for constructing such $\mathbf{Q}_0$ is given in Appendix B.

M-Schur-concave:

From definition of $\mathbf{L}$ in (25) and based Weyl' theorem [25], we have

$$[[\mathbf{L}]_{1,1}^2, \cdots, [\mathbf{L}]_{N,N}^2]^{\mathrm{T}} \prec_{\times} \sigma_b^2[\mathbf{1}_N - \boldsymbol{\lambda}(\boldsymbol{\Theta})]. \quad (33)$$

Applying $f(\bullet)$ on both sides of (33) and with **Definition 2**, we have

$$f([[\mathbf{L}]_{1,1}^2, \cdots, [\mathbf{L}]_{N,N}^2]^{\mathrm{T}}) \geq \underbrace{f(\sigma_b^2[\mathbf{1}_N - \boldsymbol{\lambda}(\boldsymbol{\Theta})])}_{\triangleq g[\boldsymbol{\lambda}(\boldsymbol{\Theta})]}. \quad (34)$$

The equality in (34) holds when $\prec_{\times}$ in (33) 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 (25), we can obtain $\sigma_b^2[\mathbf{1}_N - \boldsymbol{\lambda}(\boldsymbol{\Theta})] = [\lambda_N(\mathbf{L}\mathbf{L}^{\mathrm{H}}), \cdots, \lambda_1(\mathbf{L}\mathbf{L}^{\mathrm{H}})]^{\mathrm{T}}$. Therefore, $[[\mathbf{L}]_{1,1}^2, \cdots, [\mathbf{L}]_{N,N}^2]^{\mathrm{T}} = [\lambda_N(\mathbf{L}\mathbf{L}^{\mathrm{H}}), \cdots, \lambda_1(\mathbf{L}\mathbf{L}^{\mathrm{H}})]^{\mathrm{T}}$, which implies $\mathbf{L}$ is a diagonal matrix. With $\mathbf{L}$ being a diagonal matrix, $\mathbf{Q}_0^{\mathrm{H}}\boldsymbol{\Theta}\mathbf{Q}_0$ is also a diagonal matrix. This can be satisfied if we take $\mathbf{Q}_0 = \mathbf{U}_{\boldsymbol{\Theta}}$, where the unitary matrix $\mathbf{U}_{\boldsymbol{\Theta}}$ is defined based on the eigendecomposition $\boldsymbol{\Theta} = \mathbf{U}_{\boldsymbol{\Theta}}\boldsymbol{\Lambda}_{\boldsymbol{\Theta}}\mathbf{U}_{\boldsymbol{\Theta}}^{\mathrm{H}}$ with the elements of $\boldsymbol{\Lambda}_{\boldsymbol{\Theta}}$ arranged in decreasing order.

Notice that since $\mathbf{L}$ is a diagonal matrix, $\mathbf{C}_{\mathrm{opt}}$ in (28) is also a diagonal matrix. Based on the definition of $\mathbf{C}$ in (3) and with

the fact that $\mathbf{C}$ is a lower triangular matrix with unit diagonal elements, it can be seen that the feedback matrix $\mathbf{B}$ must be an all-zero matrix. Therefore, when the objective function is M-Schur-concave, THP becomes linear precoding. The optimality of linear transceiver for M-Schur-concave objective function has also been obtained in point-to-point MIMO systems with perfect CSI [14], [15], [17].

Remark 3: The equal bit rate assumption at the beginning of Section II is for the operation of the nonlinear precoder only (this assumption also appears in [14], [15], [19]). Notice that we have not used the equal bit rate assumption in the derivation of the optimal solution. If the objective function is chosen such that a linear transceiver is obtained, this equal bit rate assumption will not appear in the solution. On the other hand, if the objective function is chosen such that a nonlinear transceiver is obtained, the nature of the optimal transceiver is of equal bit rate (see the discussion below (32)). Therefore, the equal bit rate assumption is not a restriction.

Summary:

Summarizing the previous results, when the objective function is M-Schur-convex or M-Schur-concave, the optimization problem (29) is equivalent to

$$
\begin{aligned}
\min_{\mathbf{F}_k, \mathbf{Q}_k} \quad & g[\boldsymbol{\lambda}(\boldsymbol{\Theta})] \\
\text{s.t.} \quad & \boldsymbol{\Theta} = \mathbf{M}_1^{\mathrm{H}} \mathbf{Q}_1^{\mathrm{H}} \cdots \mathbf{M}_K^{\mathrm{H}} \mathbf{Q}_K^{\mathrm{H}} \mathbf{Q}_K \mathbf{M}_K \cdots \mathbf{Q}_1 \mathbf{M}_1 \\
& \mathrm{Tr}(\mathbf{F}_k \mathbf{F}_k^{\mathrm{H}}) \le P_k, \quad \mathbf{Q}_k^{\mathrm{H}} \mathbf{Q}_k = \mathbf{I}_{N_{R,k}}.
\end{aligned} \tag{35}
$$

where $g[\boldsymbol{\lambda}(\boldsymbol{\Theta})]$ equals to

$$
g[\boldsymbol{\lambda}(\boldsymbol{\Theta})] = \begin{cases} f(\sigma_b^2 [\prod_{n=1}^N (1 - \lambda_n(\boldsymbol{\Theta}))]^{\frac{1}{N}} \otimes \mathbf{1}_N) \\ \qquad \text{if } f(\bullet) \text{ is M-Schur-convex,} \\ f(\sigma_b^2 [\mathbf{1}_N - \boldsymbol{\lambda}(\boldsymbol{\Theta})]) \\ \qquad \text{if } f(\bullet) \text{ is M-Schur-concave.} \end{cases} \tag{36}
$$

It is difficult to directly solve the optimization problem (35), because $\boldsymbol{\Theta}$ is a product consists of matrices $\mathbf{M}_k$'s which in turn are complicated functions of the variables $\mathbf{F}_k$'s. In order to simplify the optimization problem (35), we exploit the multiplicative majorization theory and transforms the objective function of (35) to be a direct function of $\mathbf{F}_k$. To this end, we first provide useful results which form the theoretical basis of the following derivation.

B. Prerequisites of Multiplicative Majorization Theory

Definition 3: For two vectors $\mathbf{v}, \mathbf{u} \in \mathcal{D}$ with $\mathcal{D} = \{\mathbf{z} : z_1 \ge \cdots \ge z_N \ge 0\}$, $\mathbf{v} \prec_{\times, w} \mathbf{u}$ is defined as

$$
\prod_{i=1}^k v_{[i]} \le \prod_{i=1}^k u_{[i]}, \quad k = 1, \cdots, N. \tag{37}
$$

Notice that there is a subtle difference between **Definition 1** in (13) and **Definition 3**. In **Definition 3**, when $k = N$, $\prod_{i=1}^N v_{[i]} \le \prod_{i=1}^N u_{[i]}$ rather than $\prod_{i=1}^N v_{[i]} = \prod_{i=1}^N u_{[i]}$ in **Definition 1**.

Lemma 2: Let $\phi(\bullet)$ be a real-valued function on $\mathcal{D}$. Then $\phi(\bullet)$ is decreasing and multiplicatively Schur-concave on $\mathcal{D}$ if and only if

$$
\mathbf{v} \prec_{\times, w} \mathbf{u} \Rightarrow \phi(\mathbf{v}) \ge \phi(\mathbf{u}). \tag{38}
$$

Proof: See Appendix C. ■

Lemma 3: When $\phi(\bullet)$ is increasing and multiplicatively Schur-concave, for $\mathbf{v}, \mathbf{u} \in \mathcal{C} = \{\mathbf{z} : 1 > z_1 \ge \cdots \ge z_N \ge 0\}$

$$
\mathbf{v} \prec_{\times, w} \mathbf{u} \Rightarrow \phi(\mathbf{1}_N - \mathbf{v}) \ge \phi(\mathbf{1}_N - \mathbf{u}). \tag{39}
$$

Proof: See Appendix D. ■

C. Problem Reformulation

Based on the given results of multiplicative majorization theory, the optimization problem (35) 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 relationship

$$
\boldsymbol{\lambda}(\boldsymbol{\Theta}) \prec_{\times, w} \underbrace{[\gamma_1(\{\mathbf{F}_k\}_{k=1}^K), \gamma_2(\{\mathbf{F}_k\}_{k=1}^K), \cdots, \gamma_N(\{\mathbf{F}_k\}_{k=1}^K)]^{\mathrm{T}}}_{\triangleq \boldsymbol{\gamma}(\{\mathbf{F}_k\}_{k=1}^K)}
$$

with $\gamma_n(\{\mathbf{F}_k\}_{k=1}^K) = \prod_{k=1}^K \dfrac{\lambda_n(\mathbf{F}_k^{\mathrm{H}} \bar{\mathbf{H}}_k^{\mathrm{H}} \mathbf{K}_{\mathbf{F}_k}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_k)}{1 + \lambda_n(\mathbf{F}_k^{\mathrm{H}} \bar{\mathbf{H}}_k^{\mathrm{H}} \mathbf{K}_{\mathbf{F}_k}^{-1} \bar{\mathbf{H}}_k \mathbf{F}_k)}$, (40)

where the equality holds when

$$
\mathbf{Q}_k = \mathbf{V}_{\mathbf{M}_{k+1}} \mathbf{U}_{\mathbf{M}_k}^{\mathrm{H}}, \quad k = 1, \cdots, K-1 \tag{41}
$$

where $\mathbf{U}_{\mathbf{M}_k}$ and $\mathbf{V}_{\mathbf{M}_k}$ are defined based on the singular value decomposition $\mathbf{M}_k = \mathbf{U}_{\mathbf{M}_k} \boldsymbol{\Lambda}_{\mathbf{M}_k} \mathbf{V}_{\mathbf{M}_k}^{\mathrm{H}}$ with the diagonal elements of $\boldsymbol{\Lambda}_{\mathbf{M}_k}$ arranged in decreasing order. Notice that (41) does not cover the design of $\mathbf{Q}_K$, but it can be any unitary matrix because it always appears in the form $\mathbf{Q}_K^{\mathrm{H}} \mathbf{Q}_K$ and equals to an identity matrix in the objective function.

Proof: See Appendix E. ■

Property 2: The objective function $g[\boldsymbol{\lambda}(\boldsymbol{\Theta})]$ in (35) is a decreasing M-Schur-concave function with respective to $\boldsymbol{\lambda}(\boldsymbol{\Theta})$.

Proof: Based on **Lemma 2**, it is obvious that $g[\boldsymbol{\lambda}(\boldsymbol{\Theta})]$ is a decreasing M-Schur-concave function if and only if $\boldsymbol{\lambda}(\boldsymbol{\Theta}) \prec_{\times, w} \boldsymbol{\lambda}(\tilde{\boldsymbol{\Theta}}) \Rightarrow g[\boldsymbol{\lambda}(\boldsymbol{\Theta})] \ge g[\boldsymbol{\lambda}(\tilde{\boldsymbol{\Theta}})]$. In the following, we will prove the latter.

When $f(\bullet)$ is M-Schur-convex, $g[\boldsymbol{\lambda}(\boldsymbol{\Theta})] = f(\sigma_b^2 [\prod_{n=1}^N (1 - \lambda_n(\boldsymbol{\Theta}))]^{\frac{1}{N}} \otimes \mathbf{1}_N)$. Using **Lemma 1**, $\prod_{n=1}^N (1 - \lambda_n(\boldsymbol{\Theta}))$ can be proved to be a M-Schur-concave function of $\boldsymbol{\lambda}(\boldsymbol{\Theta})$. Furthermore, it can be easily seen that $\prod_{n=1}^N (1 - \lambda_n(\boldsymbol{\Theta}))$ is a decreasing function. If $\boldsymbol{\lambda}(\boldsymbol{\Theta}) \prec_{\times, w} \boldsymbol{\lambda}(\tilde{\boldsymbol{\Theta}})$ is true, based on **Lemma 2**, we have

$$
\prod_{n=1}^N (1 - \lambda_n(\boldsymbol{\Theta})) \ge \prod_{n=1}^N (1 - \lambda_n(\tilde{\boldsymbol{\Theta}})). \tag{42}
$$

Together with the fact that $f(\bullet)$ is an increasing function, it is concluded that

$$
\underbrace{f(\sigma_b^2 [\prod_{n=1}^N (1 - \lambda_n(\boldsymbol{\Theta}))]^{1/N} \otimes \mathbf{1}_N)}_{g[\boldsymbol{\lambda}(\boldsymbol{\Theta})]}
$$

$$
\ge \underbrace{f(\sigma_b^2 [\prod_{n=1}^N (1 - \lambda_n(\tilde{\boldsymbol{\Theta}}))]^{1/N} \otimes \mathbf{1}_N)}_{g[\boldsymbol{\lambda}(\tilde{\boldsymbol{\Theta}})]}. \tag{43}
$$

On the other hand, when $f(\bullet)$ is increasing and M-Schur-concave, $g[\boldsymbol{\lambda}(\boldsymbol{\Theta})] = f(\sigma_b^2[\mathbf{1}_N - \boldsymbol{\lambda}(\boldsymbol{\Theta})])$. Using **Lemma 3** we directly have $\boldsymbol{\lambda}(\boldsymbol{\Theta}) \prec_{\times,w} \boldsymbol{\lambda}(\tilde{\boldsymbol{\Theta}})$ implies

$$\underbrace{f(\sigma_b^2[\mathbf{1}_N - \boldsymbol{\lambda}(\boldsymbol{\Theta})])}_{g[\boldsymbol{\lambda}(\boldsymbol{\Theta})]} \geq \underbrace{f(\sigma_b^2[\mathbf{1}_N - \boldsymbol{\lambda}(\tilde{\boldsymbol{\Theta}})])}_{g[\boldsymbol{\lambda}(\tilde{\boldsymbol{\Theta}})]}. \qquad (44)$$

∎

Based on **Properties 1** and **2**, the objective function of (35) has an achievable lower bound $g[\boldsymbol{\lambda}(\boldsymbol{\Theta})] \geq g[\boldsymbol{\gamma}(\{\mathbf{F}_k\}_{k=1}^K)]$ with equality achieved when (41) is satisfied. When the lower bound is achieved, we have the following three additional observations:

(a) The constraints $\mathbf{Q}_k^{\mathrm{H}}\mathbf{Q}_k = \mathbf{I}_{N_{R,k}}$ 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 (41).

Applying these three observations into (35), we have the reformulated optimization problem

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

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

$$\mathrm{Tr}(\mathbf{F}_k\mathbf{F}_k^{\mathrm{H}}) \leq P_k. \qquad (45)$$

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 $\lambda_n(\mathbf{F}_k^{\mathrm{H}}\bar{\mathbf{H}}_k^{\mathrm{H}}\mathbf{K}_{\mathbf{F}_k}^{-1}\bar{\mathbf{H}}_k\mathbf{F}_k)$. Therefore, $g[\boldsymbol{\gamma}(\{\mathbf{F}_k\}_{k=1}^K)]$ is a decreasing matrix-monotone function of $\mathbf{F}_k^{\mathrm{H}}\bar{\mathbf{H}}_k^{\mathrm{H}}\mathbf{K}_{\mathbf{F}_k}^{-1}\bar{\mathbf{H}}_k\mathbf{F}_k$ [10]. Following the derivation in [11], it can be proved that at the optimal solution, the power constraints hold at the equality, i.e., $\mathrm{Tr}(\mathbf{F}_k\mathbf{F}_k^{\mathrm{H}}) = P_k$, meaning that the relays transmit at the maximum power.

Defining a variable η_{f_k} as

$$\eta_{f_k} = \alpha_k\mathrm{Tr}(\mathbf{F}_k\mathbf{F}_k^{\mathrm{H}}\boldsymbol{\Psi}_k) + \sigma_{n_k}^2 \quad \mathrm{with} \quad \alpha_k = \mathrm{Tr}(\boldsymbol{\Sigma}_k)/N_{R,k}, \qquad (46)$$

$\mathrm{Tr}(\mathbf{F}_k\mathbf{F}_k^{\mathrm{H}}) = P_k$ is exactly equivalent to $\mathrm{Tr}[\mathbf{F}_k\mathbf{F}_k^{\mathrm{H}}(\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2\mathbf{I}_{N_{T,k}})]/\eta_{f_k} = P_k$ as proved in [10], [11], [26]. Thus the robust transceiver design problem (45) is equivalent to

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

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

$$\mathrm{Tr}[\mathbf{F}_k\mathbf{F}_k^{\mathrm{H}}(\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2\mathbf{I}_{N_{T,k}})]/\eta_{f_k} = P_k. \qquad (47)$$

It is proved in Appendix F that when $\boldsymbol{\Psi}_k \propto \mathbf{I}_{N_{T,k}}$ or $\boldsymbol{\Sigma}_k \propto \mathbf{I}_{N_{R,k}}$, the optimal solutions of the optimization problem (47)

have the following structure

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

$$\mathrm{with} \quad \xi_k(\boldsymbol{\Lambda}_{\mathcal{F}_k}) = \sigma_{n_k}^2/\{1 - \alpha_k\mathrm{Tr}[\mathbf{V}_{\mathcal{H}_k,N}^{\mathrm{H}}(\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2 $$
$$\times \mathbf{I}_{N_{T,k}})^{-1/2}\boldsymbol{\Psi}_k(\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2\mathbf{I}_{N_{T,k}})^{-1/2}$$
$$\times \mathbf{V}_{\mathcal{H}_k,N}\boldsymbol{\Lambda}_{\mathcal{F}_k}^2]\}, \qquad (48)$$

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

$$(\mathbf{K}_{\mathbf{F}_k}/\eta_{f_k})^{-1/2}\bar{\mathbf{H}}_k(\alpha_k P_k \boldsymbol{\Psi}_k + \sigma_{n_k}^2\mathbf{I}_{N_{T,k}})^{-1/2}$$
$$= \mathbf{U}_{\mathcal{H}_k}\boldsymbol{\Lambda}_{\mathcal{H}_k}\mathbf{V}_{\mathcal{H}_k}^{\mathrm{H}} \qquad (49)$$

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

Remark 4: In general, the expressions of $\boldsymbol{\Psi}_k$ and $\boldsymbol{\Sigma}_k$ depend on specific channel estimation algorithms. Denote the transmit and receive antennas correlation matrices and the channel estimation error variance in the kth hop as $\mathbf{R}_{T,k}$, $\mathbf{R}_{R,k}$ and $\sigma_{e,k}^2$, respectively. When the channels are estimated based on the algorithm proposed in [21], [22], it can be shown that $\boldsymbol{\Psi}_k = \mathbf{R}_{T,k}$ and $\boldsymbol{\Sigma}_k = \sigma_{e,k}^2(\mathbf{I}_{N_{R,k}} + \sigma_{e,k}^2\mathbf{R}_{R,k}^{-1})^{-1}$. If the transmit antennas or the receive antennas are spaced widely, we have $\mathbf{R}_{T,k} \propto \mathbf{I}_{N_{T,k}}$ or $\mathbf{R}_{R,k} \propto \mathbf{I}_{N_{R,k}}$. These imply $\boldsymbol{\Psi}_k \propto \mathbf{I}_{N_{T,k}}$ or $\boldsymbol{\Sigma}_k \propto \mathbf{I}_{N_{R,k}}$. Moreover, if the length of training is large, the value of $\sigma_{e,k}^2$ will be small and $\mathbf{I}_{N_{R,k}} + \sigma_{e,k}^2\mathbf{R}_{R,k}^{-1} \approx \mathbf{I}_{N_{R,k}}$. As a result, $\boldsymbol{\Sigma}_k$ will also approximate an identity matrix even when $\mathbf{R}_{R,k} \not\propto \mathbf{I}_{N_{R,k}}$. On the other hand, if the channel statistics are unknown, and using least-square channel estimator, it can be derived that $\boldsymbol{\Sigma}_k \propto \mathbf{I}_{N_{R,k}}$ always holds regardless of the antenna correlation or training length [8].

B. Computation of $\boldsymbol{\Lambda}_{\mathcal{F}_k}$

It is obvious that in (48), the only unknown variable is $\boldsymbol{\Lambda}_{\mathcal{F}_k}$. In the following, we will discuss how to solve $\boldsymbol{\Lambda}_{\mathcal{F}_k}$ in more detail. Denoting the following diagonal elements as

$$[\boldsymbol{\Lambda}_{\mathcal{H}_k}]_{n,n} = h_{k,n}, \qquad [\boldsymbol{\Lambda}_{\mathcal{F}_k}]_{n,n} = f_{k,n}, \qquad (50)$$

substituting (48) into the optimization problem (47) and noticing that $\xi_k(\boldsymbol{\Lambda}_{\mathcal{F}_k}) = \eta_{f_k}$ (shown by (79) in Appendix F), after a straightforward derivation, the optimization for robust transceiver design is simplified as

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

$$\mathrm{s.t.} \quad \gamma_n(\{\mathbf{F}_k\}_{k=1}^K) = \prod_{k=1}^K \frac{f_{k,n}^2 h_{k,n}^2}{f_{k,n}^2 h_{k,n}^2 + 1}$$

$$\sum_{n=1}^N f_{k,n}^2 = P_k. \qquad (51)$$

The solution of (51) depends on whether $f(\bullet)$ is M-Schur-convex or M-Schur-concave.

M-Schur-convex functions:

Notice that when $f([\text{MSE}_1, \cdots, \text{MSE}_N]^\text{T})$ is an M-Schur-convex function, regardless of the specific expression of $f(\bullet)$, the optimization problem (51) is equivalent to minimize $\prod_{n=1}^{N}(1 - \gamma_n(\{\mathbf{F}_k\}_{k=1}^K))$ [15]. Therefore, the transceiver design problem (51) equals to

$$\min_{f_{k,i}} \quad \sum_{n=1}^{N} \log \left(1 - \frac{\prod_{k=1}^{K} f_{k,n}^2 h_{k,n}^2}{\prod_{k=1}^{K}(f_{k,n}^2 h_{k,n}^2 + 1)} \right)$$

$$\text{s.t.} \quad \sum_{n=1}^{N} f_{k,n}^2 = P_k. \tag{52}$$

In order to solve the optimization problem (52), iterative water-filling 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

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

$$\left. -1 \right)^+ \qquad n = 1, \cdots, N$$

$$\text{with} \quad a_{k,n} = \prod_{l \neq k} f_{l,n}^2 h_{l,n}^2/(f_{l,n}^2 h_{l,n}^2 + 1) \tag{53}$$

where μ_k is the Lagrange multiplier which makes $\sum_{n=1}^{N} f_{k,n}^2 = P_k$ hold [27]. Notice that this iterative water-filling algorithm is guaranteed to converge, as discussed in [28].

M-Schur-concave functions:

When $f([\text{MSE}_1, \cdots, \text{MSE}_N]^\text{T})$ is a M-Schur-concave functions, there is no unified solution. In this case, $\mathbf{\Lambda}_{\mathcal{F}_k}$ should be solved case by case. In the following, we use the example $f([\text{MSE}_1, \cdots, \text{MSE}_N]^\text{T}) = \prod_{n=1}^{N} \text{MSE}_n^{w_n}$ for $w_1 \geq w_2 \cdots \geq w_N \geq 0$ to illustrate how to compute $\mathbf{\Lambda}_{\mathcal{F}_k}$. For this objective function, using (36) it follows that $g[\mathbf{\gamma}(\{\mathbf{F}_k\}_{k=1}^K)] = \sigma_b^{2\sum_n w_n} \prod_{n=1}^{N} (1 - \gamma_n(\{\mathbf{F}_k\}_{k=1}^K))^{w_n}$ and the optimization (51) is equivalent to

$$\min_{f_{k,i}} \quad \sum_{n=1}^{N} w_n \log \left(1 - \frac{\prod_{k=1}^{K} f_{k,n}^2 h_{k,n}^2}{\prod_{k=1}^{K}(f_{k,n}^2 h_{k,n}^2 + 1)} \right)$$

$$\text{s.t.} \quad \sum_{n=1}^{N} f_{k,n}^2 = P_k. \tag{54}$$

Equation (54) has the same form as (52). Therefore, the solution can also be obtained by iterative water-filling solution. Notice that the design problem becomes linear transceiver design problem when $f(\bullet)$ is M-Schur-concave.

C. Summary and Implementation Issues

The design idea and procedure of the proposed robust transceiver are summarized in Table I. For the implementation of the proposed algorithm, the execution order is in reverse, i.e., from Step 8 to Step 3. Notice that in Step 8, iterative water-filling is adopted to solve for $\mathbf{\Lambda}_{\mathcal{F}_k}$. In general, only local optimality of the solution can be guaranteed [29], which is a common problem for AF MIMO relaying design [6], [9], [26].

TABLE I
SUMMARY OF ROBUST TRANSCEIVER DESIGN

1.	Derive the data estimation MSE matrix (10).
2.	Formulate the optimization problem (12) with $\mathbf{G}$, $\mathbf{P}_k$ and $\mathbf{C}$ as variables.
3.	Derive the optimal equalizer $\mathbf{G}$ as a function of $\mathbf{P}_k$ and $\mathbf{C}$ given by (15). Substitute the optimal $\mathbf{G}$ into the optimization problem (12) to reduce the number of variables and have a reformulated optimization problem (18).
4.	Simplify the constraints of the optimization problem (18) by replacing $\mathbf{P}_k$ with $\mathbf{Q}_k$ and $\mathbf{F}_k$, and obtain an equivalent optimization problem (24).
5.	Derive the optimal $\mathbf{C}$ as a function of $\mathbf{Q}_k$ and $\mathbf{F}_k$ given by (28). Substitute the optimal $\mathbf{C}$ into the optimization problem (24) and reformulate the optimization problem as (29).
6.	Derive the optimal $\mathbf{Q}_k$ as a function of $\mathbf{F}_k$ based on majorization theory and substitute the optimal $\mathbf{Q}_k$ in (41) into the optimization problem (29) to reduce the number of variables. The optimization problem is then simplified to be (45).
7.	Derive the optimal structure of $\mathbf{F}_k$ given by (48).
8.	Solve for the unknown diagonal matrices $\mathbf{\Lambda}_{\mathcal{F}_k}$ in the optimal structure using (51).

For information sharing in the implementation of the proposed solution, we can consider two algorithms.

Central Algorithm:

In centralized implementation, a natural assumption is that there is a central node performing the transceiver designs. All other nodes send its own estimated CSI to the central node via control channels, and after completing the design the central node informs each node the corresponding transceiver matrix. Since the channel does not change (or change very slowly), estimated CSI transmitted on control channels can be considered error-free due to low data transmission rates and heavy channel coding.

Distributed Algorithm:

Based on the derived optimal structure $\mathbf{F}_{k,\text{opt}}$ in (48) and (49) and the optimal $\mathbf{Q}_k$ in (41), using the definition of $\mathbf{F}_k$ given by (19) and (20), we can derive that the optimal forwarding matrix at the kth node has the following structure

$$\mathbf{P}_k = (\alpha_k P_k \mathbf{\Psi}_k + \sigma_{n_k}^2 \mathbf{I}_{N_{T,k}})^{-1/2}$$
$$\times \mathbf{V}_{\mathcal{H}_k, N} \mathbf{\Lambda}_{\mathbf{P}_k} \mathbf{U}_{\mathcal{H}_{k-1}, N}^\text{H} \mathbf{K}_{\mathbf{F}_{k-1}}^{-1/2}, \tag{55}$$

where $\mathbf{\Lambda}_{\mathbf{P}_k}$ is a diagonal matrix whose elements are functions of the diagonal elements of $\mathbf{\Lambda}_{\mathcal{F}_m}$ for all m. It can be seen that except $\mathbf{\Lambda}_{\mathbf{P}_k}$, all other matrices in (55) are only the functions of the channels immediately preceding and succeeding the kth node. It is easy for each node to obtain such channel information. As a result, the only information shared among all the other nodes is the diagonal elements of matrix $\mathbf{\Lambda}_{\mathcal{F}_m}$ denoted by $\{f_{m,n}\}_{n=1}^N$. Notice that $\{f_{m,n}\}_{n=1}^N$ is the solution of the optimization problem (51). Exploiting the linear network topology and the fact that in the first constraint of (51) $\{f_{m,n}\}_{n=1}^N$ appears in the form of $\prod_{k=1}^{K} f_{k,n}^2 h_{k,n}^2/(f_{k,n}^2 h_{k,n}^2 + 1)$, only local information needs to be shared between adjacent nodes.

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
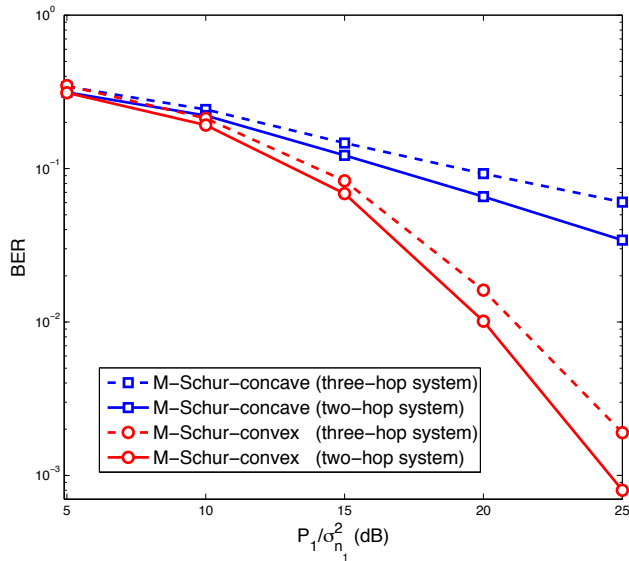
Fig. 2. BERs of the proposed transceivers with M-Schur-convex and M-Schur-concave objective functions when $\rho_t = 0$, $\rho_r = 0.4$, $\sigma_e^2 = 0.001$.
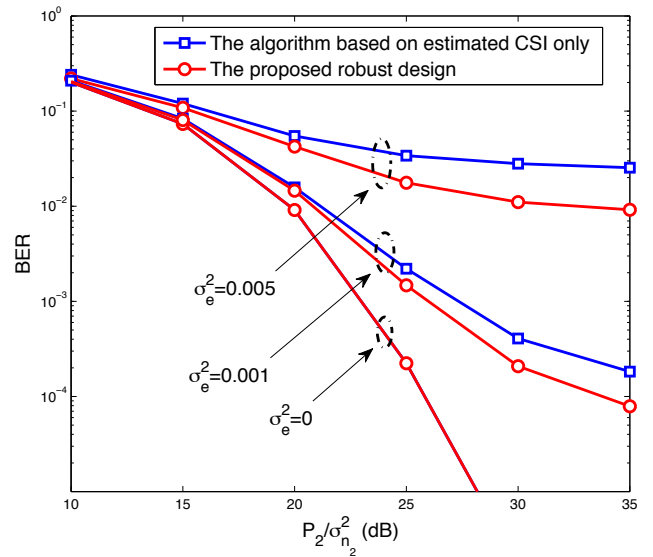

Fig. 3. BERs of proposed robust design with M-Schur-convex objective functions and the algorithm based on estimated CSI only when $\rho_t = 0.5$, $\rho_r = 0$, and $P_1/\sigma_{n_1}^2 = 30$dB.

and destination are all equipped with four antennas, i.e., $N_{T,k} = N_{R,k} = 4$. The estimation error correlation matrices are chosen as the popular exponential model $[\boldsymbol{\Psi}_k] = \sigma_e^2 \rho_t^{|i-j|}$ and $[\boldsymbol{\Sigma}_k] = \rho_r^{|i-j|}$ [7] 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 [7], [30], [31]

$$\bar{\mathbf{H}}_k \sim \mathcal{CN}_{N_{R,k}, N_{T,k}}\left(\mathbf{0}_{N_{R,k}, N_{T,k}}, \frac{(1-\sigma_e^2)}{\sigma_e^2} \boldsymbol{\Sigma}_k \otimes \boldsymbol{\Psi}_k^{\mathrm{T}}\right), \quad (56)$$

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, $N_{\mathrm{Data}} = 10000$ independent 16-QAM symbols are transmitted. Each point in the following figures is an average of 10000 trials and the bit error rates (BER) are computed [33]–[35].

First, we consider the objective function as weighted geometric mean MSE with equal weighting. In this case, the objective function is both M-Schur-convex and M-Schur-concave. There are two optimal solutions: one being linear transceiver and the other one being nonlinear transceiver. Fig. 2 compares the BERs of these two solutions. Both two-hop and three-hop systems are simulated with $\rho_t = 0$, $\rho_r = 0.4$, $\sigma_e^2 = 0.001$, $P_2/\sigma_{n_2}^2 = P_3/\sigma_{n_3}^2 = 30$dB and $P_1/\sigma_{n_1}^2$ being varied from 5 to 30dB. As expected, the nonlinear transceiver has a better performance than linear transceiver, but the performance improvement of nonlinear transceiver comes at the expense of higher complexity. Comparing to linear transceiver, the THP nonlinear transceiver has an additional $N \times N$ triangular matrix multiplication. Thus the additional complexity is $N(1 + N)/2$ complex multiplications and $N(N-1)/2$ complex additions for each vector transmission. Furthermore, although the three-hop system performs not as good as the two-hop system, due to the extra hop of channel and noise amplification, the performance of the two-hop and three-hop systems shows the same trend. In the following,

we focus on the M-Schur-convex objective function (i.e., nonlinear transceiver) for two-hop system only.

Next, we investigate the effect of the channel estimation error on the BER performance. Fig. 3 shows the BERs of the proposed robust nonlinear design and the corresponding algorithm based on estimated CSI only (which takes the channel estimates as true channels) with $\rho_t = 0.5$, $\rho_t = 0$, $P_1/\sigma_{n_1}^2 = 30$dB, and $P_2/\sigma_{n_2}^2$ being varied from 10 to 35dB. The algorithm based on estimated CSI only is obtained by simply setting $\boldsymbol{\Psi}_k = \mathbf{0}$ in the proposed algorithm (similar approach has been used in [20] and [21]). From Fig. 3, 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. Furthermore, the performance gap between the proposed robust design and the algorithm based on estimated CSI becomes larger as the channel estimation error increases. Of course, the performance of the two algorithms coincide when $\sigma_e^2 = 0$.

Finally, we illustrate the effects of correlation in the channel estimation errors. Fig. 4 shows the BERs of the proposed robust design with M-Schur-convex objective functions and the corresponding algorithm based estimated CSI only for different ρ_r, when $\rho_t = 0$, $\sigma_e^2 = 0.002$, $P_1/\sigma_{n_1}^2 = 30$dB, and $P_2/\sigma_{n_2}^2$ being varied from 10 to 35dB. It can be seen that in addition to the fact that the performance of the proposed robust design is always better than that based on the estimated CSI only, as ρ_r increases, the performance gain of the proposed robust design with respect to that based on CSI only becomes larger. It is most obvious when $\rho_r = 0.9$ and at high SNR at the second hop. The performance gaps come from the fact that when correlation becomes stronger, $\boldsymbol{\Sigma}_k$ will be very different from identity matrix. Therefore from (48) and (49), the proposed optimal structure will be significantly different from that of the algorithm with estimated CSI only. As the designed precoding and forwarding matrices can be considered
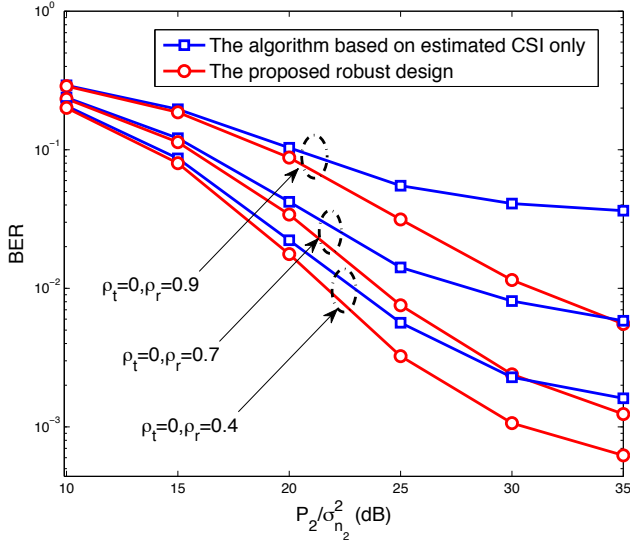
Fig. 4. BERs of proposed robust design with M-Schur-convex objective functions and the algorithm based on estimated CSI only with different ρ_r, when $\rho_t = 0$, $\sigma_e^2 = 0.002$ and $P_1/\sigma_{n_1}^2 = 30$dB.
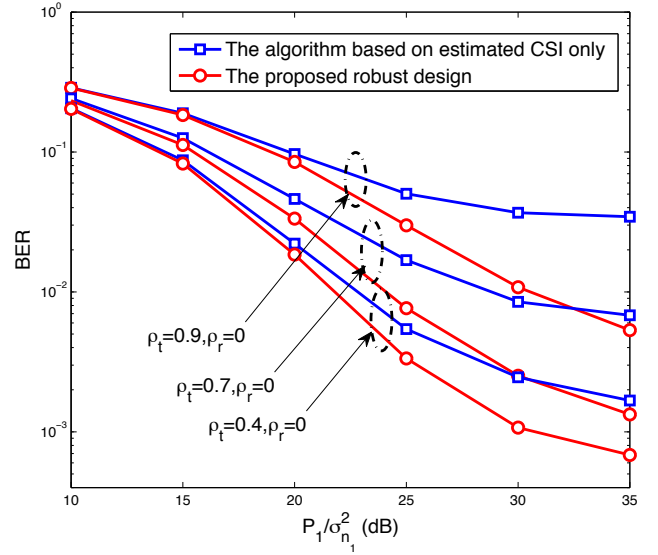

Fig. 5. BERs of proposed robust design with M-Schur-convex objective functions and the algorithm based on estimated CSI only with different ρ_t, when $\rho_r = 0$, $\sigma_e^2 = 0.002$ and $P_2/\sigma_{n_2}^2 = 30$dB.

as the transmission directions, Fig. 4 shows that correlation of channel estimation error would affect the direction of data transmission, and subsequently affect the final BER performance. Fig. 5 shows the corresponding BERs for different ρ_t, with $\rho_r = 0$, $\sigma_e^2 = 0.002$, $P_2/\sigma_{n_2}^2 = 30$dB, and $P_1/\sigma_{n_1}^2$ being varied from 10 to 35dB. It can be seen that a similar conclusion can be drawn.

VII. CONCLUSIONS

Joint Bayesian robust transceiver design for multi-hop AF MIMO relaying systems was investigated. It was assumed that channel estimation errors exist in CSI in all hops. At the source node, a nonlinear Tomlinson-Harashima precoding 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-Schur-convex or M-Schur-concave. Using 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 LEMMA 1

Based on **Definition 2**, $\phi(\mathbf{z})$ is M-Schur-convex over $\mathcal{D} = \{\mathbf{z} : z_1 \geq \cdots \geq z_N \geq 0\}$ if and only if for $\mathbf{v}, \mathbf{u} \in \mathcal{D}$, $\mathbf{v} \prec_{\times} \mathbf{u}$ implies $\phi(\mathbf{v}) \leq \phi(\mathbf{u})$.

For a vector $\mathbf{z} \in \mathcal{D}$, define

$$\tilde{\mathbf{z}} = [\tilde{z}_1, \cdots, \tilde{z}_N]^{\mathrm{T}} \quad \text{and} \quad \tilde{z}_k = \prod_{i=1}^{k} z_i. \tag{57}$$

For $\mathbf{v}, \mathbf{u} \in \mathcal{D}$, it is obvious that $\mathbf{v} \prec_{\times} \mathbf{u}$ is equivalent to

$$\{\tilde{v}_k \leq \tilde{u}_k\}_{k=1}^{N-1}, \quad \text{and} \quad \tilde{v}_N = \tilde{u}_N. \tag{58}$$

On the other hand, based on (57), z_k equals to

$$z_k = \tilde{z}_k / \tilde{z}_{k-1}, \quad k \leq L_z, \tag{59}$$

where $L_z - 1$ is the number of the nonzero elements of $\mathbf{z}$. Therefore $\phi(\mathbf{v}) \leq \phi(\mathbf{u})$ can be written as

$$\underbrace{\phi(\tilde{v}_1, \tilde{v}_2/\tilde{v}_1 \cdots, \tilde{v}_{L_v}/\tilde{v}_{L_v-1}, 0, \cdots,)}_{\triangleq \psi(\tilde{\mathbf{v}})}$$
$$\leq \underbrace{\phi(\tilde{u}_1, \tilde{u}_2/\tilde{u}_1 \cdots, \tilde{u}_{L_u}/\tilde{u}_{L_u-1}, 0, \cdots,)}_{\triangleq \psi(\tilde{\mathbf{u}})}, \tag{60}$$

Based on (58) and (60), proving $\phi(\mathbf{z})$ is M-Schur-convex is equivalent to proving when $\{\tilde{v}_k \leq \tilde{u}_k\}_{k=1}^{N-1}$ and $\tilde{v}_N = \tilde{u}_N$ hold, we have $\psi(\tilde{\mathbf{v}}) \leq \psi(\tilde{\mathbf{u}})$. In other words, the proof becomes to prove $\psi(\bullet)$ is a vector-valued increasing function.

To prove $\psi(\bullet)$ is increasing, we only need to prove that when $\tilde{v}_k \leq \tilde{u}_k$ and $\tilde{v}_l = \tilde{u}_l$ for all $l \neq k$, we have $\psi(\tilde{\mathbf{v}}) \leq \psi(\tilde{\mathbf{u}})$ [24]. As $\tilde{v}_k \geq 0$ and $\tilde{u}_k \geq 0$, $\tilde{v}_k \leq \tilde{u}_k$ is equivalent to $\tilde{v}_k = \tilde{u}_k/e$ with $e \geq 1$. Substituting $\tilde{v}_k = \tilde{u}_k/e$ and $\tilde{v}_l = \tilde{u}_l$ for all $l \neq k$ into (60) and replacing $u_k = \tilde{u}_k/\tilde{u}_{k-1}$ for $k \leq L_u - 1$, proving $\psi(\tilde{\mathbf{v}}) \leq \psi(\tilde{\mathbf{u}})$ is equivalent to proving $\phi(u_1, \cdots, u_k/e, u_{k+1}e, \cdots)$ is increasing over $e \geq 1$ and $u_k/e \geq u_{k+1}e$.

APPENDIX B
ALGORITHM FOR COMPUTING $\mathbf{Q}_0$

Following the sufficient conditions given in [32], an explicit algorithm for constructing $\mathbf{Q}_0$ is given as follows. Without loss of generality, in this Appendix, for both singular value decomposition (SVD) and eigendecomposition, the elements of the diagonal singular value or eigenvalue matrix are assumed to be in decreasing order.

Step 1: Define $\mathbf{A}$ based on the following eigen-decomposition

$$(\mathbf{I}_N - \mathbf{M}_1^{\mathrm{H}}\mathbf{Q}_1^{\mathrm{H}}\cdots\mathbf{M}_K^{\mathrm{H}}\mathbf{Q}_K^{\mathrm{H}}\mathbf{Q}_K\mathbf{M}_K\cdots\mathbf{Q}_1\mathbf{M}_1)^{1/2}\sigma_b$$
$$= \underbrace{\mathbf{U_M}\mathbf{\Lambda_M}\mathbf{U_M^H}}_{\triangleq\mathbf{A}}. \qquad (61)$$

Step 2: Initialize $\mathbf{S} = \mathbf{0}_{N\times N}$ and set

$$[\mathbf{S}]_{1,1} = \sqrt{\frac{|\mathbf{A}^{\mathrm{H}}\mathbf{A}|^{1/N} - [\mathbf{\Lambda_M}]_{N,N}}{[\mathbf{\Lambda_M}]_{1,1} - [\mathbf{\Lambda_M}]_{N,N}}},$$
$$[\mathbf{S}]_{N,1} = \sqrt{\frac{[\mathbf{\Lambda_M}]_{1,1} - |\mathbf{A}^{\mathrm{H}}\mathbf{A}|^{1/N}}{[\mathbf{\Lambda_M}]_{1,1} - [\mathbf{\Lambda_M}]_{N,N}}}. \qquad (62)$$

Meanwhile, the orthogonal complement matrix of $[\mathbf{S}]_{:,1}$ is set to be

$$[\mathbf{S}]_{:,1}^{\perp} = \begin{bmatrix} -[\mathbf{S}]_{1,1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \\ [\mathbf{S}]_{N,1} & \mathbf{0} \end{bmatrix}. \qquad (63)$$

Step 3: Begin recursion for $k = 1,\cdots,N-2$. Compute a $(N-k)\times(N-k)$ unitary matrix $\mathbf{V}^{(k)}$ based on the following eigendecomposition

$$(\mathbf{A}[\mathbf{S}]_{:,1:k}^{\perp})^{\mathrm{H}}[\mathbf{I} - \mathbf{A}[\mathbf{S}]_{:,1:k}([\mathbf{S}]_{:,1:k}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}\mathbf{A}[\mathbf{S}]_{:,1:k})^{-1}$$
$$\times [\mathbf{S}]_{:,1:k}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}](\mathbf{A}[\mathbf{S}]_{:,1:k}^{\perp}) = \mathbf{V}^{(k)}\mathbf{\Lambda}^{(k)}(\mathbf{V}^{(k)})^{\mathrm{H}}. \qquad (64)$$

Then update the $(k+1)^{\mathrm{th}}$ column of $\mathbf{S}$ as

$$[\mathbf{S}]_{:,k+1} = [\mathbf{S}]_{:,1:k}^{\perp}\mathbf{V}^{(k)}\mathbf{y}^{(k)} \qquad (65)$$

and

$$\mathbf{y}^{(k)} = \left[\sqrt{\frac{|\mathbf{A}^{\mathrm{H}}\mathbf{A}|^{1/N} - [\mathbf{\Lambda}^{(k)}]_{N-k,N-k}}{[\mathbf{\Lambda}^{(k)}]_{1,1} - [\mathbf{\Lambda}^{(k)}]_{N-k,N-k}}}, \mathbf{0}_{1,N-k-1},\right.$$
$$\left.\sqrt{\frac{[\mathbf{\Lambda}^{(k)}]_{1,1} - |\mathbf{A}^{\mathrm{H}}\mathbf{A}|^{1/N}}{[\mathbf{\Lambda}^{(k)}]_{1,1} - [\mathbf{\Lambda}^{(k)}]_{N-k,N-k}}}\right]^{\mathrm{T}}. \qquad (66)$$

Based on the SVD $\mathbf{S} = \mathbf{U_S}\mathbf{\Lambda_S}\mathbf{V_S^H}$, the orthogonal complement matrix of $[\mathbf{S}]_{:,1:k+1}$ is computed as

$$[\mathbf{S}]_{:,1:k+1}^{\perp} = [\mathbf{U_S}]_{:,k+2:N}. \qquad (67)$$

Step 4: When $k = N-1$, $[\mathbf{S}]_{:,N} = [\mathbf{S}]_{:,1:N-2}^{\perp}\mathbf{V}^{(N-2)}\mathbf{y}^{(N-1)}$ and

$$\mathbf{y}^{(N-1)} = \left[\sqrt{\frac{[\mathbf{\Lambda}^{(N-2)}]_{1,1} - |\mathbf{A}^{\mathrm{H}}\mathbf{A}|^{1/N}}{[\mathbf{\Lambda}^{(N-2)}]_{1,1} - [\mathbf{\Lambda}^{(N-2)}]_{2,2}}},\right.$$
$$\left.-\sqrt{\frac{|\mathbf{A}^{\mathrm{H}}\mathbf{A}|^{1/N} - [\mathbf{\Lambda}^{(N-2)}]_{2,2}}{[\mathbf{\Lambda}^{(N-2)}]_{1,1} - [\mathbf{\Lambda}^{(N-2)}]_{2,2}}}\right]^{\mathrm{T}}. \qquad (68)$$

Step 5: Finally, $\mathbf{Q}_0$ equals to $\mathbf{Q}_0 = \mathbf{U_M}\mathbf{S}$.

APPENDIX C
PROOF OF LEMMA 2

Proof of "if" direction

First, we will prove that for any two vectors $\mathbf{v},\mathbf{u} \in \mathcal{D}$, $\mathbf{v} \prec_{\times,w} \mathbf{u} \Rightarrow \phi(\mathbf{v}) \geq \phi(\mathbf{u})$ implies $\phi(\bullet)$ is a decreasing M-Schur-concave function over $\mathcal{D}$.

When $\mathbf{v} \prec_{\times,w} \mathbf{u} \Rightarrow \phi(\mathbf{v}) \geq \phi(\mathbf{u})$ holds, $\mathbf{v} \prec_{\times} \mathbf{u} \Rightarrow \phi(\mathbf{v}) \geq \phi(\mathbf{u})$ must hold. Using **Lemma 1**, $\phi(\bullet)$ must be M-Schur-concave over $\mathcal{D}$.

Furthermore, for $\mathbf{v},\mathbf{u} \in \mathcal{D}$ with $v_k \leq u_k$ and $v_i = u_i$ for all $i \neq k$, we have $\mathbf{v} \prec_{\times,w} \mathbf{u}$. Then $\mathbf{v} \prec_{\times,w} \mathbf{u} \Rightarrow \phi(\mathbf{v}) \geq \phi(\mathbf{u})$ implies $\phi(\bullet)$ is a decreasing function. Therefore, when $\mathbf{v} \prec_{\times,w} \mathbf{u} \Rightarrow \phi(\mathbf{v}) \geq \phi(\mathbf{u})$, then we have $\phi(\bullet)$ is a decreasing M-Schur-concave function.

Proof of "only if" direction

On the other hand, when $\phi(\bullet)$ is a decreasing M-Schur-concave function, we need prove that $\mathbf{v} \prec_{\times,w} \mathbf{u} \Rightarrow \phi(\mathbf{v}) \geq \phi(\mathbf{u})$. For any two vectors $\mathbf{v},\mathbf{u} \in \mathcal{D}$ with $\mathbf{v} \prec_{\times,w} \mathbf{u}$ we can construct a vector $\boldsymbol{\tau} \in \mathcal{D}$ with $\tau_i = u_i$ for $i < N$ and τ_N is chosen to makes $\prod_{i=1}^{N}\tau_i = \prod_{i=1}^{N}v_i$. It is obvious that $\tau_N \leq u_N$. Then if $\mathbf{v} \prec_{\times,w} \mathbf{u}$, we have $\mathbf{v} \prec_{\times} \boldsymbol{\tau}$ and $\boldsymbol{\tau} \prec_{\times,w} \mathbf{u}$.

As $\phi(\bullet)$ is M-Schur-concave, based on **Lemma 1** we directly have $\phi(\mathbf{v}) \geq \phi(\boldsymbol{\tau})$. Furthermore, since the difference between $\boldsymbol{\tau}$ and $\mathbf{u}$ is only in the last element with $\tau_N \leq u_N$, as $\phi(\bullet)$ is decreasing, we have $\phi(\boldsymbol{\tau}) \geq \phi(\mathbf{u})$. Combining the two inequalities, we have $\phi(\mathbf{v}) \leq \phi(\mathbf{u})$.

APPENDIX D
PROOF OF LEMMA 3

Based on **Lemma 1**, it can be proved that $\prod_{i=1}^{k}(1-z_i)$ is an M-Schur-concave function. It is also obvious that $\prod_{i=1}^{k}(1-z_i)$ is a decreasing function for $\mathbf{z} \in \mathcal{C} = \{\mathbf{z} : 1 > z_1 \geq \cdots \geq z_N \geq 0\}$. Using **Lemma 2**, for $\mathbf{v},\mathbf{u} \in \mathcal{C}$ with $\mathbf{v} \prec_{\times,w} \mathbf{u}$, we have

$$\prod_{i=1}^{k}\underbrace{(1-v_i)}_{\triangleq\hat{v}_{(i)}} \geq \prod_{i=1}^{k}\underbrace{(1-u_i)}_{\triangleq\hat{u}_{(i)}} > 0, \quad k = 1,\cdots,N. \qquad (69)$$

We construct a vector $\hat{\boldsymbol{\tau}} = [\hat{\tau}_{(1)},\cdots,\hat{\tau}_{(N)}]^{\mathrm{T}}$ with $\hat{\tau}_{(i)} = \hat{u}_{(i)}$ for $i < N$ and $\hat{\tau}_{(N)}$ is chosen to makes $\prod_{i=1}^{N}\hat{v}_{(i)} = \prod_{i=1}^{N}\hat{\tau}_{(i)}$ hold. It is obvious that $\tau_{(N)} \geq u_{(N)}$. As the only difference between $\tau_{(i)}$ and $u_{(i)}$ is at $i = N$, when $\phi(\bullet)$ is increasing, we have $\phi(\hat{\boldsymbol{\tau}}) \geq \phi(\hat{\mathbf{u}})$ where $\hat{\mathbf{u}} = [\hat{u}_{(1)},\cdots,\hat{u}_{(N)}]^{\mathrm{T}}$.

On the other hand, based on (69) and the fact that $\hat{\tau}_{(i)} = \hat{u}_{(i)}$ for $i < N$, it can be concluded that (a) $\prod_{i=1}^{k}\hat{v}_{(i)} \geq \prod_{i=1}^{k}\hat{\tau}_{(i)}$ for $1 \leq k < N$. Based on the definition of $\hat{\tau}_N$, it can also be concluded that (b) $\prod_{i=1}^{N}\hat{v}_{(i)} = \prod_{i=1}^{N}\hat{\tau}_{(i)} > 0$. Results (a) and (b) implies $\hat{\mathbf{v}} \prec_{\times} \hat{\boldsymbol{\tau}}$ where $\hat{\mathbf{v}} = [\hat{v}_{(1)},\cdots,\hat{v}_{(N)}]^{\mathrm{T}}$ [24]. As $\phi(\bullet)$ M-Schur-concave, using **Lemma 1**, we have $\phi(\hat{\mathbf{v}}) \geq \phi(\hat{\boldsymbol{\tau}})$. Together with the conclusion in the last paragraph, we can obtain $\phi(\hat{\mathbf{v}}) \geq \phi(\hat{\mathbf{u}})$. Finally, with $\hat{\mathbf{v}} = \mathbf{1}_N - \mathbf{v}$ and $\hat{\mathbf{u}} = \mathbf{1}_N - \mathbf{u}$, the proof is completed.

APPENDIX E
PROOF OF PROPERTY 1

First notice two facts in matrix theory: (a) for two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ with compatible dimension $\lambda_i(\boldsymbol{AB}) = \lambda_i(\boldsymbol{BA})$ [24, 9.A.1.a]; (b) for two positive semi-definite matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, $\prod_{i=1}^{n}\lambda_i(\boldsymbol{AB}) \leq \prod_{i=1}^{n}\lambda_i(\boldsymbol{A})\lambda_i(\boldsymbol{B})$ [24, 9.H.1.a], where the equality holds when $\boldsymbol{A}$ and $\boldsymbol{B}$ has the same unitary matrix in

eigendecomposition. With these two facts, we have

$$
\prod_{i=1}^{n} \lambda_i(\mathbf{M}_1^{\mathrm{H}}\mathbf{Q}_1^{\mathrm{H}}\cdots\mathbf{M}_K^{\mathrm{H}}\mathbf{Q}_K^{\mathrm{H}}\mathbf{Q}_K\mathbf{M}_K\cdots\mathbf{Q}_1\mathbf{M}_1)
$$
$$
= \prod_{i=1}^{n} \lambda_i(\mathbf{M}_2^{\mathrm{H}}\mathbf{Q}_2^{\mathrm{H}}\cdots\mathbf{M}_K^{\mathrm{H}}\mathbf{Q}_K^{\mathrm{H}}\mathbf{Q}_K\mathbf{M}_K\cdots
$$
$$
\times\, \mathbf{Q}_2\mathbf{M}_2\mathbf{Q}_1\mathbf{M}_1\mathbf{M}_1^{\mathrm{H}}\mathbf{Q}_1^{\mathrm{H}})
$$
$$
\leq \prod_{i=1}^{n} [\lambda_i(\mathbf{M}_2^{\mathrm{H}}\mathbf{Q}_2^{\mathrm{H}}\cdots\mathbf{M}_K^{\mathrm{H}}\mathbf{Q}_K^{\mathrm{H}}\mathbf{Q}_K\mathbf{M}_K\cdots\mathbf{Q}_2\mathbf{M}_2)
$$
$$
\times\, \lambda_i(\mathbf{M}_1\mathbf{M}_1^{\mathrm{H}})] \qquad n = 1,\cdots,N, \tag{70}
$$

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

$$
\prod_{i=1}^{n} \lambda_i(\boldsymbol{\Theta})
$$
$$
\leq \prod_{i=1}^{n} \underbrace{[\lambda_i(\mathbf{M}_K^{\mathrm{H}}\mathbf{M}_K)\lambda_i(\mathbf{M}_{K-1}^{\mathrm{H}}\mathbf{M}_{K-1})\cdots\lambda_i(\mathbf{M}_1^{\mathrm{H}}\mathbf{M}_1)]}_{\triangleq \gamma_i(\{\mathbf{F}_k\}_{k=1}^{K})},
\tag{71}
$$

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

$$
\mathbf{Q}_k = \mathbf{V}_{\mathbf{M}_{k+1}}\mathbf{U}_{\mathbf{M}_k}^{\mathrm{H}}, \quad k = 1,\cdots,K-1, \tag{72}
$$

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 (23), $\gamma_i(\{\mathbf{F}_k\}_{k=1}^{K})$ in (71) equals to

$$
\gamma_i(\{\mathbf{F}_k\}_{k=1}^{K}) = \prod_{k=1}^{K} \frac{\lambda_i(\mathbf{F}_k^{\mathrm{H}}\bar{\mathbf{H}}_k^{\mathrm{H}}\mathbf{K}_{\mathbf{F}_k}^{-1}\bar{\mathbf{H}}_k\mathbf{F}_k)}{1 + \lambda_i(\mathbf{F}_k^{\mathrm{H}}\bar{\mathbf{H}}_k^{\mathrm{H}}\mathbf{K}_{\mathbf{F}_k}^{-1}\bar{\mathbf{H}}_k\mathbf{F}_k)}. \tag{73}
$$

APPENDIX F
OPTIMAL STRUCTURE OF $\mathbf{F}_k$

Defining new variables

$$
\tilde{\mathbf{F}}_k = 1/\sqrt{\eta_{f_k}}(\alpha_k P_k\boldsymbol{\Psi}_k + \sigma_{n_k}^2\mathbf{I}_{N_{T,k}})^{1/2}\mathbf{F}_k, \quad \text{and}
$$
$$
\boldsymbol{\mathcal{H}}_k = (\mathbf{K}_{\mathbf{F}_k}/\eta_{f_k})^{-1/2}\bar{\mathbf{H}}_k(\alpha_k P_k\boldsymbol{\Psi}_k + \sigma_{n_k}^2\mathbf{I}_{N_{T,k}})^{-1/2}, \tag{74}
$$

the optimization problem (47) is reformulated as

$$
\min_{\tilde{\mathbf{F}}_k} \quad g[\boldsymbol{\gamma}(\{\tilde{\mathbf{F}}_k\}_{k=1}^{K})]
$$
$$
\text{s.t.} \quad \gamma_n(\{\tilde{\mathbf{F}}_k\}_{k=1}^{K}) = \prod_{k=1}^{K} \frac{\lambda_n(\tilde{\mathbf{F}}_k^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k\tilde{\mathbf{F}}_k)}{1 + \lambda_n(\tilde{\mathbf{F}}_k^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k\tilde{\mathbf{F}}_k)}
$$
$$
\mathrm{Tr}(\tilde{\mathbf{F}}_k\tilde{\mathbf{F}}_k^{\mathrm{H}}) = P_k. \tag{75}
$$

When $\boldsymbol{\Psi}_k \propto \mathbf{I}_{N_{T,k}}$ or $\boldsymbol{\Sigma}_k \propto \mathbf{I}_{N_{R,k}}$, for the optimal solution $\mathbf{K}_{\mathbf{F}_k}/\eta_{f_k}$ is constant [10], [11], [26] and thus $\boldsymbol{\mathcal{H}}_k$ is constant. Let $\tilde{\mathbf{F}}_{k,\mathrm{opt}}$ be the optimal solution of (75). With the following singular value decompositions,

$$
\boldsymbol{\mathcal{H}}_k\tilde{\mathbf{F}}_{k,\mathrm{opt}} = \mathbf{U}_{A_k}\boldsymbol{\Lambda}_{A_k}\mathbf{V}_{A_k}^{\mathrm{H}}, \quad \boldsymbol{\mathcal{H}}_k = \mathbf{U}_{\mathcal{H}_k}\boldsymbol{\Lambda}_{\mathcal{H}_k}\mathbf{V}_{\mathcal{H}_k}^{\mathrm{H}}, \tag{76}
$$

where the diagonal elements of $\boldsymbol{\Lambda}_{A_k}$ and $\boldsymbol{\Lambda}_{\mathcal{H}_k}$ are arranged in decreasing order, we can construct a matrix $\hat{\mathbf{F}}_k$ equals to

$$
\hat{\mathbf{F}}_k = \mathbf{V}_{\mathcal{H}_k}\boldsymbol{\Lambda}_{\mathbf{X}_k}\mathbf{V}_{A_k}^{\mathrm{H}}, \tag{77}
$$

where $\boldsymbol{\Lambda}_{\mathbf{X}_k}$ is a rectangular diagonal matrix with the same rank as $\boldsymbol{\Lambda}_{A_k}$ and $1/b_k\boldsymbol{\Lambda}_{\mathcal{H}_k}\boldsymbol{\Lambda}_{\mathbf{X}_k} = \boldsymbol{\Lambda}_{A_k}$. The scalar b_k is chosen to make that $\mathrm{Tr}(\hat{\mathbf{F}}_k\hat{\mathbf{F}}_k^{\mathrm{H}}) = P_k$ holds.

Using **Lemma 12** in [23], we can show that $\hat{\mathbf{F}}_k^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k\hat{\mathbf{F}}_k \succeq \tilde{\mathbf{F}}_{k,\mathrm{opt}}^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k\tilde{\mathbf{F}}_{k,\mathrm{opt}}$. Together with the formulation of $\gamma_n(\{\tilde{\mathbf{F}}_k\}_{k=1}^{K})$ in (75), it can be concluded that $\gamma_n(\{\hat{\mathbf{F}}_k\}_{k=1}^{K}) \geq \gamma_n(\{\tilde{\mathbf{F}}_{k,\mathrm{opt}}\}_{k=1}^{K})$. Since $g(\bullet)$ is an decreasing function, $g[\boldsymbol{\gamma}(\{\hat{\mathbf{F}}_k\}_{k=1}^{K})] \leq g[\boldsymbol{\gamma}(\{\tilde{\mathbf{F}}_{k,\mathrm{opt}}\}_{k=1}^{K})]$. Because $\tilde{\mathbf{F}}_{k,\mathrm{opt}}$ is the optimal solution, it is impossible to have $g[\boldsymbol{\gamma}(\{\hat{\mathbf{F}}_k\}_{k=1}^{K})] < g[\boldsymbol{\gamma}(\{\tilde{\mathbf{F}}_{k,\mathrm{opt}}\}_{k=1}^{K})]$. Therefore, $\hat{\mathbf{F}}_k$ must be the optimal solution. Furthermore, based on the relationship between of $\tilde{\mathbf{F}}_k$ and $\mathbf{F}_k$, it follows that

$$
\mathbf{F}_{k,\mathrm{opt}} = \sqrt{\eta_{f_k}}(\alpha_k P_k\boldsymbol{\Psi}_k + \sigma_{n_k}^2\mathbf{I}_{N_{T,k}})^{-1/2}\mathbf{V}_{\mathcal{H}_k}\boldsymbol{\Lambda}_{\mathbf{X}_k}\mathbf{V}_{A_k}^{\mathrm{H}}. \tag{78}
$$

Notice that in general the unitary matrix $\mathbf{V}_{A_k}$ depends on the optimal solution $\tilde{\mathbf{F}}_{k,\mathrm{opt}}$. However, from (75), it can be seen that the value of $\mathbf{V}_{A_k}$ does not affect the objective functions and therefore it can be an arbitrary unitary matrix. Meanwhile, as the minimum dimension of $\tilde{\mathbf{F}}_k^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k^{\mathrm{H}}\boldsymbol{\mathcal{H}}_k\tilde{\mathbf{F}}_k$ is N, only $N\times N$ principal submatrix of $\boldsymbol{\Lambda}_{\mathbf{X}_k}$ can be nonzero. For notational convenience, we denote that $[\boldsymbol{\Lambda}_{\mathbf{X}_k}]_{1:N,1:N} = \boldsymbol{\Lambda}_{\boldsymbol{\mathcal{F}}_k}$.

Substituting (78) into the definition of η_{f_k} in (46), we obtain a simple linear function of η_{f_k}, and η_{f_k} can be easily solved to be

$$
\eta_{f_k} = \sigma_{n_k}^2/\{1 - \alpha_k\mathrm{Tr}[\mathbf{V}_{\mathcal{H}_k,N}^{\mathrm{H}}(\alpha_k P_k\boldsymbol{\Psi}_k + \sigma_{n_k}^2\mathbf{I}_{N_{T,k}})^{-1/2}
$$
$$
\times\, \boldsymbol{\Psi}_k(\alpha_k P_k\boldsymbol{\Psi}_k + \sigma_{n_k}^2\mathbf{I}_{N_{T,k}})^{-1/2}\mathbf{V}_{\mathcal{H}_k,N}\boldsymbol{\Lambda}_{\boldsymbol{\mathcal{F}}_k}^2]\}
$$
$$
\triangleq \xi_k(\boldsymbol{\Lambda}_{\boldsymbol{\mathcal{F}}_k}). \tag{79}
$$

REFERENCES

[1] A. Swami, Q. Zhao, Y.-W. Hong, and L. Tong, *Wireless Sensor Networks: Signal Processing and Communications*, Wiley Press, 2007.

[2] 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.

[3] 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.

[4] 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.

[5] F.-S. Tseng, W.-R. Wu, and J.-Y. Wu "Joint source/relay precoder design in nonregenerative cooperative systems using an MMSE criterion," *IEEE Trans. Wireless Commun.*, vol. 8, no. 10, pp. 4928–4933, Oct. 2009.

[6] 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.

[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, 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.

[9] 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.

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

[11] C. Xing, S. Ma, Z. Fei, Y.-C. Wu, and J. Kuang, "Joint robust weighted LMMSE transceiver design for dual-hop AF multiple-antenna relay systems," in *Proc. IEEE Globecom 2011*, Houston, TX, USA, Dec. 2011.

[12] 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.

[13] 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.

[14] M. B. Shenouda and T. N. Davidson, "A framework for designing MIMO systems with decision feedback equalization or Tomlinson-Harashima precoding," *IEEE J. Sel. Areas Commun.*, vol. 26 no. 2 pp.401–411, Feb. 2008.

[15] 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.

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

[17] D. P. Palomar and Y. Jiang, "MIMO Transceiver Design via Majorization Theory," *Foundations and Trends in Communications and Information Theory*, Now Publishers, vol. 3, no. 4, pp. 331–551, Nov. 2006.

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

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

[20] 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.

[21] 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.

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

[23] 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. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, Sep. 2003.

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

[25] H. Weyl, "Inequality between the two kinds of eigenvalues of a linear transformation," *Proc. Nat. Acad. Sci. USA*, vol. 35, no. 7, pp. 408–411, July 1949.

[26] C. Xing, Z. Fei, S. Ma, J. Kuang, and Y.-C. Wu, "Robust linear transceiver design for multi-hop non-regenerative MIMO relaying systems," in *Proc. IEEE WCSP 2011*, Nanjing, China, Nov. 2011.

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

[28] W. Yu, W. Rhee, S. Boyd, and J. Cioffi, "Iterative water-filling for Gaussian vector multiple access channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp.145–152, Jan. 2004.

[29] 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

[30] L. Musavian, M. R. Nakhi, M. Dohler, 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.

[31] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inf. Theory* vol. 52, no. 5, pp. 2203–2214, May 2006.

[32] J.-K. Zhang, A. Kavčić, and K. M. Wong, "Equal-diagonal QR decomposition and its application to precoder design for successive-cancellation detection," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 154–172, Jan. 2005.

[33] H. Zhu and J. Wang, "Chunk-Based Resource Allocation in OFDMA Systems-Part I Chunk Allocation," *IEEE Trans. Commun.*, vol. 57, no. 9, pp. 2734–2744, Sep. 2009.

[34] Y. Zhou, J. Wang, T.S. Ng, K. Higuchi, and M. Sawahashi, "OFCDM: a promising broadband wireless access technique," *IEEE Commun. Magazine*, vol. 46, no.3, pp. 39–49, March 2008.

[35] Y. Zhou, J. Wang, and M. Sawahashi, "Downlink transmission of broadband OFCDM systems–Part I: Hybrid Detection," *IEEE Trans. Commun.*, vol. 53, no. 4, pp. 718-729, April 2005.

Chengwen Xing (S'08–M'10) received the B.Eng. degree from Xidian University, Xi'an, China, in 2005 and the Ph.D. degree in the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, in 2010. Since September 2010, he has been with the School of Information and Electronics, Beijing Institute of Technology (BIT), Beijing, China, where he is currently a Lecturer. He was a TPC member for ICC 2012 and Globecom 2012. His current research interests include statistical signal processing, convex optimization, multivariate statistics, combinatorial optimization, MIMO systems, and cooperative communication systems.

Minghua Xia (M'12) obtained his Ph.D. degree in Telecommunications and Information Systems from Sun Yat-sen University, Guangzhou, China, in 2007.

From March 2007 to July 2009, he was with the Electronics and Telecommunications Research Institute (ETRI) of South Korea, Beijing R&D Center, Beijing, China, where he worked as a member of engineering staff and participated in the projects on the physical layer design of 3GPP LTE mobile communications. From Aug. 2009 to Feb. 2011, he was with The University of Hong Kong (HKU), Hong Kong, as a Postdoctoral Fellow. Currently, he is with King Abdullah University of Science and Technology (KAUST), Saudi Arabia. His research interests are in the area of network information theory and space-time signal processing, and in particular the design and performance analysis of multi-user multi-antenna systems, cooperative relaying systems, and cognitive radio networks.

Feifei Gao (S'05–M'09) received the Ph.D. degree from National University of Singapore, Singapore, in 2007. He was a Research Fellow with the Institute for Infocomm Research, A*STAR, Singapore, in 2008 and was an Assistant Professor with the School of Engineering and Science, Jacobs University, Bremen, Germany, from 2009 to 2010. In 2011, he joined the Department of Automation, Tsinghua University, Beijing, China, where he is currently an Associate Professor. He also serves as an Adjunct Professor with the School of Engineering and Science, Jacobs University Bremen, Germany.

Prof. Gao has authored and coauthored more than 100 refereed IEEE journal and conference proceeding papers. He has served as an editor for IEEE Wireless Communications Letters and as a Technical Program Committee member for the IEEE International Conference on Communications, the IEEE Global Communications Conference, the IEEE Vehicular Technology Conference, and the IEEE Personal, Indoor, and Mobile Radio Communications Conference.

Yik-Chung Wu received the B.Eng. (EEE) degree in 1998 and the M.Phil. degree in 2001 from the University of Hong Kong (HKU). He received the Ph.D. degree from Texas A&M University, College Station, in 2005. From August 2005 to August 2006, he was with the Thomson Corporate Research, Princeton, NJ, as a Member of Technical Staff. Since September 2006, he has been with HKU as an Assistant Professor. He was a visiting scholar at Princeton University in summer 2011. His research interests are in general area of signal processing and communication systems. He is currently serving as an editor for the IEEE Communications Letters, and Journal of Communications and Networks.